

Identification of molecular genetic causes of bull infertility

Dissertation

to obtain the Ph. D. degree

in the International Ph. D. Program for Agricultural Sciences in Göttingen (IPAG)

at the Faculty of Agricultural Sciences,

Georg-August-University Göttingen, Germany



presented by

Shuwen Shan

Born in Nei Mongol, China

Göttingen, June 2021

D7

1. Referee: Prof. Prof. h.c. mult. Dr. Dr. Bertram Brenig

2. Referee: Prof. Dr. Jens Tetens

Date of dissertation: 02.06.2021

Table of Contents

Table of Contents	i
List of publications.....	ii
Abstract	iii
Abbreviations	v
CHAPTER 1 General Introduction.....	1
<i>1 Bull fertility.....</i>	<i>2</i>
1.1 The phenotype of bull fertility	2
1.1.1 Overview	2
1.1.2 Assessment and indicator	2
1.1.3 Infertility and subfertility	4
1.2 Molecular biological research of male fertility in cattle.....	4
1.2.1 Related QTL regions and genes	4
1.2.2 Epigenetic study	6
1.2.3 Causative variants for bull sub- or infertility	7
<i>2 Overview of hemophilia B Leyden</i>	<i>7</i>
<i>3 Gene function of Kruppel-like factor 7 (KLF7)</i>	<i>8</i>
CHAPTER 2 Association of α/β-Hydrolase D16B with Bovine Conception Rate and Sperm Plasma Membrane Lipid Composition.....	15
CHAPTER 3 Christmas disease in a Hovawart family resembling human hemophilia B Leyden is caused by a single nucleotide deletion in a highly conserved transcription factor binding site of the <i>F9</i> gene promoter.....	34
CHAPTER 4 A Missense Mutation in the <i>KLF7</i> Gene Is a Potential Candidate Variant for Congenital Deafness in Australian Stumpy Tail Cattle Dogs	52
CHAPTER 5 General Discussion	67
<i>1 Significance of the research.....</i>	<i>68</i>
<i>2 Biological studies in animals help humans understand themselves.....</i>	<i>68</i>
<i>3 Candidate Gene Targeting</i>	<i>69</i>
3.1 Phenotype definition and measurement.....	69
3.2 GWAS and WGS.....	70
<i>4 Post-GWAS era</i>	<i>71</i>
4.1 Annotation of candidate variants and genes	71
4.2 Functional verification of candidate variants and genes.....	73
4.3 Application and perspective.....	74
Acknowledgments	vii
Curriculum Vitae.....	x

List of publications

This thesis is based on the first three papers.

1. **Shan, S.[†]**, Xu, F.[†], Bleyer, M., Becker, S., Melbaum, T., Wemheuer, W., Hirschfeld, M., Wacker, C., Zhao, S., Schütz, E. and Brenig, B., 2020. Association of α/β -Hydrolase D16B with Bovine Conception Rate and Sperm Plasma Membrane Lipid Composition. *International Journal of Molecular Sciences*, 21(2), p.627.
2. Brenig, B.[†], Steingraber, L.[†], **Shan, S.**, Xu, F., Hirschfeld, M., Andag, R., Spengeler, M., Dietschi, E., Mischke, R. and Leeb, T., 2019. Christmas disease in a Hovawart family resembling human hemophilia B Leyden is caused by a single nucleotide deletion in a highly conserved transcription factor binding site of the F9 gene promoter. *Haematologica*, 104(11), p.2307.
3. Xu, F.[†], **Shan, S.[†]**, Sommerlad, S., Seddon, J.M. and Brenig, B., 2021. A Missense Mutation in the KLF7 Gene Is a Potential Candidate Variant for Congenital Deafness in Australian Stumpy Tail Cattle Dogs. *Genes*, 12(4), p.467.
4. **Shan, S.**, Xu, F. and Brenig, B., 2021. Genome-Wide Association Studies Reveal Neurological Genes for Dog Herding, Predation, Temperament, and Trainability Traits. *Frontiers in veterinary science*, 8.
5. **Shan, S.**, Xu, F., Hirschfeld, M., and Brenig, B., 2021. Sperm Lipid Markers of Male Fertility in Mammals. *International Journal of Molecular Sciences*, 22(16), p.8767.
6. **Shan, S.**, Xu, F., Bleyer, M., Becker, S., Melbaum, T., Wemheuer, W., Hirschfeld, M., Wacker, C., Zhao, S., Schütz, E. and Brenig, B., 2019. Molecular genetic analysis of male infertility in Holstein Friesian cattle. 37th *International Society for Animal Genetics*, 7-12 July 2019, Lleida.
7. **Shan, S.**, Xu, F., Bleyer, M., Becker, S., Melbaum, T., Wemheuer, W., Hirschfeld, M., Wacker, C., Zhao, S., Schütz, E. and Brenig, B., 2019. Tarantino's infertility. *DGfZ-Jahrestagung und DGfZ-/GfT-Gemeinschaftstagung 2019*. 11-12 September 2019, Gießen.
8. **Shan, S.**, Xu, F., Sommerlad, S., Seddon and Brenig, B., 2021. KLF7 gene is a risk factor for congenital deafness in Australian Stumpy Tail Cattle Dogs. 38th *International Society for Animal Genetics*, 26-30 July 2021, Virtual.

[†] These authors have equally contributed.

Abstract

This is a cumulative thesis that includes three different topics around male infertility in cattle and canine diseases. Whole genome association analysis (GWAS) and whole genome sequencing (WGS), as well as additional experimental approaches, were used to identify the candidate genes for these disorders.

One of the current challenges facing the dairy industry is to predict and improve the fertility of bulls. We have detected a Holstein bull, who had been approved for artificial insemination based on his semen characteristics, but did not produce offspring after 412 first inseminations, resulting in a non-return rate (NR-Abw) of -29. To identify causative mutations underlying this idiopathic infertility, GWAS and WGS were performed on different non-return rate bulls followed by genotyping of candidate variants in a large population. Finally, a nonsense mutation (AC_000170.1: g.54429815G>A, rs468948776) in α/β -Hydrolase D16B (ABHD16B) on chromosome 13 was identified as a potential candidate variant. Protein analysis revealed expression of ABHD16B in the testes and epididymis of control bulls, and lipidomic analysis showed significant differences in lipid content of the plasma membrane between carriers and controls. We conclude that *ABHD16B* may play a role in lipid changes during spermatogenesis or sperm maturation and that the altered lipid content may explain the reduced fertilization capacity of mutant sperm.

Hemophilia B is a monogenic, X-chromosome recessive bleeding disorder caused by genetic variants in the coagulation factor IX gene (*F9*). Here we identify a Hovawart family with hemophilia B Leyden caused by a deletion (NC_006621.3:g.109501492delC) in the *F9* promoter within the conserved binding region of hepatocyte nuclear factor 4 α (HNF-4 α) and androgen receptor (AR). Using EMSA we found that the deletion only eliminated the binding of HNF-4 α to the promoter, but did not affect the binding of AR. *In vitro* approach utilizing luciferase reporter gene assay revealed a significant expression reduction of the mutated promoter compared to wild-type. Our results suggest that dogs provide a natural occurrence model for understanding hemophilia B.

Congenital deafness is prevalent in many modern dog breeds, but few causative genes have been identified to date. Here, we attempt to find the genetic cause of congenital deafness around three deaf Australian Stumpy Tail Cattle Dogs (ASCD) using GWAS and WGS. A

GWAS was performed on 3 bilateral deaf ASCDs, 43 herding dogs, and one unaffected ASCD, resulting in the identification of 13 significantly associated loci on 6 chromosomes (CFA3, 8, 17, 23, 28, and 37). More than half of the significantly associated signals were localized on CFA37, containing the most significantly associated variant. Further WGS was conducted on same three deaf ASCDs and one control, whereupon a missense variant (NC_006619.3:g.15562684G>A; XP_022270984.1:p.Leu173Phe) of kruppel-like factor 7 (*KLF7*) on CFA 37 was considered as a candidate variant. Genotyping of the *KLF7* variant in an additional cohort of 55 ASCD samples (28 deaf and 27 normal hearing dogs) indicated that the variant was still associated with deafness ($p = 0.014$), with a calculated penetrance of 0.75. In conclusion, *KLF7* is a promising candidate gene for the etiology of deafness in ASCD.

Abbreviations

3C	Chromosome conformation capture
4C	Chromosome conformation capture-on-chip
5C	Chromosome conformation capture carbon copy
AAV	Adeno-associated viruses
ABHD16B	α/β -hydrolase domain-containing 16B
ADAM28	Disintegrin and metalloproteinase domain-containing protein 28
AI	Artificial insemination
AR	Acrosome reaction
AR	Androgen receptor
ARMC3	Armadillo repeat-containing protein 3
ASCD	Australian stumpy tail cattle dog
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
BSE	Breeding soundness evaluation
BTA	Bovine chromosome
C/EBP α	CCAAT/enhancing-binding protein α
CCDC189	Coiled-coil domain containing 189
CCT6A	T-complex protein 1 subunit zeta
CCT8	T-complex protein 1 subunit theta
CD163	CD163 Molecule
ChIA-PET	Chromatin interaction analysis by paired-end tag sequencing
ChIP	Chromatin immunoprecipitation
CKB	Creatine kinase B
CRISP2	Cysteine rich secretory protein 2
DDX4	DEAD-box helicase 4
DNA	Deoxyribonucleic acid
DNAJA1	DnaJ homolog subfamily A member 1
ERCR	Estimated relative conception rate
FAANG	Functional Annotation of ANimal Genomes
FBXO38	F-box only protein 38
FGF2	Fibroblast growth factor 2
FSHB	Follicle-stimulating hormone beta-subunit
GATA3	GATA binding protein 3
GO	Gene Ontology
GTE _x	Genotype-Tissue Expression
GWAS	Genome-wide association study
HNF4 α	Hepatocyte nuclear factor 4 α
IGF1R	Insulin-like growth factor 1
ITGB5	Integrin beta 5
KAT8	Lysine acetyltransferase 8
KLFs	Kruppel-like factors
MAP1B	Microtubule-associated protein 1B

miRNA	MicroRNA
mRNA	Messenger RNA
NRR	Non-return rate
ONECUT1/2	One cut homeobox
PEBP1	Phosphatidylethanolamine-binding protein 1
PIWIL3	Piwi-like protein 3
PPP1R11	Protein phosphatase 1 regulatory subunit 11
QTL	Quantitative trait loci
RNA	Ribonucleic Acid
ROH	Runs of homozygosity
scATAC-seq	Single-cell ATAC-seq
SCR	Sire conception rate
SNP	Single nucleotide polymorphism
SPO11	SPO11 Initiator of meiotic double stranded breaks
STAT5A	Signal transducer and activator of transcription 5A
TBC1D20	TBC1 domain family member 20
TCaRE3	TRKB Ca ²⁺ Response Element 3
TDRD9	Tudor domain containing 9
TMEM119	Transmembrane protein 119
TMEM95	Transmembrane protein 95
TRKB	Tropomyosin receptor kinase B
TSS	Transcriptional start site
UKLF	Ubiquitous kruppel-like factor
WGS	Whole-genome sequencing
ZFN	Zinc-finger nucleases

CHAPTER 1

General Introduction

1 Bull fertility

1.1 The phenotype of bull fertility

1.1.1 Overview

Fertility is an important economic parameter in livestock industry [1, 2]. Cattle are significant sources of milk and meat products for human. With the increased population and the growing economy, people have been pursuing high-yielding cattle breeding to meet the growing demand. However, the transfer to more productive cows is related to a decline in reproductive performance [3]. Actually, reproductive efficiency is also an important factor for dairy producers to pursue economic maximization. Increased reproductive capacity can raise the economic net returns of entire herds [4]. Therefore, many researchers began to explore the balance between seeking higher yield and maintaining reproductive performance [5]. In addition, many other factors should be considered to explain the deterioration of cattle's reproductive status, such as nutrition, disease, management, inbreeding and environment (reviewed by Lucy [5]).

The cattle genetic pool is continuously shrinking due to a high attention to limited economically important traits and the widespread use of a selected number of elite sires via artificial insemination (AI). Caraviello et al. [6] conducted a comprehensive survey of management practices on reproductive performance and obtained data from 103 herds. As expected, all of these herds predominantly used AI sires. In cattle breeding, most studies have focused on the fertility traits of cows, but ignored the role of bulls. Recently, more and more studies have appealed to increase the attention on bull fertility [7].

1.1.2 Assessment and indicator

The potential fertility of bulls can be assessed by physical examination and phenotypic assessments. The fertility of bulls is affected by various factors, and no single criterion or diagnostic test can assess it reliably. Therefore, several measurements or indicators have been developed that are usually combined to predict the bull fertility.

In terms of physical examination, it is generally believed that testicular characteristics and sperm parameters are closely related to bull fertility [8, 9]. Scrotal circumference is commonly used because it is a good indicator of sperm output. Scrotal size has positive genetic correlations with testis size and weight, sperm quantity [10] and sperm morphology [11]. Sperm parameters, especially compensable traits (i.e., motility and morphology), are

employed in routine assessment of bull fertility, although these are only moderately correlated with fertility and show large ranges of variation [12]. Sperm motility is one of the most important characteristics associated with the fertilizing ability of bulls [13-17]. Some evidences suggest that abnormal sperm morphology is correlated with poor fertility [17-19]. However, the accuracy of morphological parameters in fertility prediction depends on the degree and percentage of abnormal sperm morphology [20, 21]. A breeding soundness evaluation (BSE) is the physical and reproductive examination for testicular development, sperm motility and morphology, which is the routine procedure for removal of severely abnormal bulls in respect of reproductive value from the herd. Some evidences suggest that implementation of bull BSE can improve the fertility and production performance of cattle [22, 23]

In addition, the noncompensable traits and physiological functions of sperm are crucial for bull fertility evaluation. Noncompensable traits are those traits that cannot be overcome by increasing the number of spermatozoa in insemination [24]. Physiological functions are essential for spermatozoa to complete the specific process in fertilization (reviewed by Rodriguez-Martinez [25]). The defects of those traits may not manifest until spermatozoa-oocyte binding, fusion or early embryo development. For example, the spermatozoic chromatin structure was correlated with bull fertility [26]. It is reported that the proportion of non-capacitated spermatozoa of frozen-thawed semen was positively correlated to fertility [27]. The state of acrosome reaction (AR) or the ability of sperm-zona pellucida binding is significantly correlated with non-return rate (NRR) [15, 28]. Several studies also demonstrated the paternal critical effect on developmental kinetics [29], blastocyst yield and quality [30, 31].

However, laboratory analysis alone is not sufficient to fully evaluate the fertility of sires. The bull's accurate breeding records can provide phenotypic assessments of their own fertility. Several fertility endpoints between the time of insemination and the birth of offspring can estimate fertility including fertilization rate, NRR, and conception rate. Fertilization rate is the ability of sperm from a particular individual to fertilize an egg. An *in vivo* assessment of fertilization rate is flushing the uterus on day 6 or 7 after insemination to obtain the unfertilized ova or embryos of varying developmental stages from cows [32]. NRR, the percentage of cows that are not subsequently re-bred in a given interval after an insemination, has been used in bull fertility evaluation for a long time due to its easily detectable properties

[32, 33]. A 70-day NRR-based bull fertility assessment termed the estimated relative conception rate (ERCR) dates back to 1986. Since 2008, sire conception rate (SCR) based on confirmed pregnancy has begun to evaluate fertility of AI service-sires [34, 35]. In fact, these indicators from field data cannot assess 100% bull fertility, because cows, microenvironment and herd management have a combined impact on the outcomes [33, 36].

1.1.3 Infertility and subfertility

Infertility in humans is defined as a couple failing to induce a pregnancy within 1 year of regular unprotected intercourse [37]. According to a study in 2015, the distribution of infertility caused by male factors ranges from 20% to 70%, and the proportion of infertile men ranges from 2.5% to 12% [38]. At least 15% of these infertile men are associated with genetic disorders that include chromosomal and single-gene aberrations [39]. Similarly, bovine infertility is a complex problem, manifested by decreased fertility rates, low conception rates and reduced embryo survival ability [5, 40].

Male infertility is a complex multifactorial disease with highly heterogeneous phenotype. It can be associated with congenital or acquired factors acting at pretesticular, testicular, and post-testicular levels [41]. The causes of male infertility vary widely including the disorders of hypothalamic–pituitary axis function, quantitative and qualitative problems of spermatogenesis, and ductal obstruction or dysfunction [42]. Among them, quantitative spermatogenesis defects caused by primary testicular failure, i.e. from azoospermia to oligozoospermia, are the most common cause (about 75%) [41, 43]. Besides, age, general and systemic diseases, the toxin and environmental factors as well as psychological factors all play important roles in male infertility. In addition, there is a large number of infertility cases of at least 30% with unknown causes, known as ‘idiopathic infertility’ [37].

In cattle, many defects and causes of infertility have been summarized [44, 45]. Chenoweth [46] has reviewed some genetic defects of sperm, which is the main reason for bull infertility. Furthermore, the pathogens that cause bull infertility or the pathogens that could be transmitted by bovine semen were reviewed by Givens and Marley [47].

1.2 Molecular biological research of male fertility in cattle

1.2.1 Related QTL regions and genes

The growing application of molecular biological techniques in bovine studies has facilitated

the discovery of complex mechanisms of reduced male fertility. With the advancement of high-throughput technology and the availability of large amounts of data, genome-wide association study (GWAS), genome-wide scanning and other methods have been successfully applied for fertility estimation. Many regions and genes have been assumed to be indicators of bull fertility and the latest review summarized recent achievements [48].

Some molecular markers for testicular size, semen characteristics or sperm functions of bulls have been developed. A research of puberty in Brahman bulls illustrated that genomic regions on bovine chromosome 2 (BTA2), BTA14, and chromosome X are associated with hormone levels, scrotal circumference and normal sperm percentage [49]. Some GWAS have declared a number of single nucleotide polymorphism (SNP) markers, candidate genes and regions that are significantly associated with semen production traits, e.g., semen volume, number of sperms, motility and morphometry [50-53]. Feugang et al. [54] identified 97 markers that were significantly associated with fertility, in which integrin beta 5 (*ITGB5*) gene was confirmed to be involved in fertilization. The application of targeted sequencing of fertility-related β -defensin genes and whole exome sequencing of AI bulls with varying fertility phenotypes has revealed many fertility-associated SNPs. The haplotype harboring several β -defensin genes is highly correlated with the density of sperm binding to oviductal epithelial explants [55]. The SNP associations with noncompensatory fertility, which cannot be overcome by increasing the number of sperms, were estimated with 795 Holstein sires using a weighted Bayes B analysis, and 14 chromosomes regions showed different degree of signals. However, based on the low heritability of noncompensatory fertility, the accuracy of this prediction was low [56].

There are also some molecular markers associated with the fertility evaluation of field data phenotypes. A correlation analysis showed that *FGF2* and *STAT5A* polymorphisms were associated with ERCR in a population of 222 Holstein bulls [57]. Peñagaricano et al. [58] conducted a GWAS and identified eight SNPs that were significantly associated with SCR. Some SNPs are located near or inside genes related to fertility, and some SNPs show marked dominant effects. Subsequently, an association analysis between highly conserved spermatogenesis genes from the fly to human and SCR was analyzed in an US Holstein population, and 5 significantly related SNPs located in three genes (*MAP1B*, *PPP1R11*, *DDX4*) were detected [59]. A comprehensive genomic study has illustrated eight genomic regions (BTA5, BTA9, BTA13, BTA21, and BTA25) related to Holstein bull fertility using

SCR as a parameter [60]. In these regions, many genes are associated with sperm physiology and male fertility. For example, *CKB* [61], *TDRD9* [62], *IGF1R* [63], *KAT8* [64] and *CCT6A* [65] are involved in sperm biology such as spermatogenesis, sperm capacitation and sperm-oocyte interaction. A study revealed the non-additive effects on SCR of Holstein bulls. Four genomic regions on BTA8, BTA9, BTA13 and BTA17 exhibited dominance and/or recessive effects by using a two-step mixed model-based approach [66]. Most regions contain genes related to bull fertility, such as *ADAM28* [67, 68], *DNAJAI* [69], *TBC1D20* [70], *SPO11* [71], *PIWIL3* [72] and *TMEM119* [73], which play roles in testicular development, spermatogenesis and sperm maturation. In addition, a study on US Jersey dairy cattle identified ten regions on eight chromosomes that can account for more than 0.5% of the additive genetic variance for SCR. When non-additive effect was considered, two regions on BTA11 and BTA25 exhibited recessive effects. Most regions harbor genes with known functions of bull fertility [74]. A recent study detected eight runs of homozygosity (ROH) regions that were significantly associated with SCR, and the candidate genes involved in these regions were directly related to male fertility [75].

1.2.2 Epigenetic study

In addition to the contribution of DNA, spermatozoa also provide RNAs, transcription factors and cell signaling molecules during the fertilization process [76]. For instance, Kropp et al. [77] tried to reveal the RNA expression and epigenetic characteristics of spermatozoa and embryos between bulls with different fertility status. 98 differentially expressed genes were detected in embryos which were generated by high or low fertility bulls, and 76 differentially methylated regions were found in the spermatozoa of these bulls. Lalancette et al. [78] demonstrated that sperms in high-NRR group displayed a higher proportion of transcripts in metabolism, signal transduction, translation, glycosylation, and protein degradation compared to low-NRR group. Arangasamy et al. [79] found an association between mRNA expression of *CRISP2*, *CCT8* and *PEBP1* in spermatozoa and different SCR scores in Holstein bulls. Card et al. [80] attempted to use RNA-sequencing to compare the oligo-dT selected sperm transcript profiles of bulls with varying fertility. Govindaraju et al. [81] have illustrated some differentially expressed miRNAs in the sperm from high and low fertility bulls and suggested that the miRNAs may play important roles in bull fertility. In addition, as more data are generated, DNA methylation information is being used as a biomarker of fertility in bulls. Gross et al. [82] integrated DNA methylation data and RNA-sequencing data between high-

and low-fertility sires, identified 1765 differentially methylated cytosines and determined ten genes as candidate markers of bull fertility.

A cohort of protein markers of fertility has also been identified using bulls with varying fertility. Peddinti et al. [83] detected 125 potential protein biomarkers that are differentially expressed in high and low fertility bull spermatozoa. D'Amours et al. [84] reported eight differentially expressed proteins in high and low fertility bulls, some of which are related to the proportion of live sperm. Furthermore, some sperm related proteins were detected, e.g., seminal plasma proteins [85], fertility-associated antigen [86] and the protein P25b on sperm [87], which can be used as potential indicators.

1.2.3 Causative variants for bull sub- or infertility

Although many biomarkers and genes related to bull fertility have been announced, only a few have been identified as causative mutations for cattle subfertility or infertility. A mutation in the 5'-upstream regulatory regions of *FSHB* can lead to decreased serum follicle-stimulating hormone levels and further result in poor semen quality and low fertility in bulls [88, 89]. Another variant related to poor semen quality has been reported in Swedish Red cattle. A frameshift mutation in *ARMC3* can cause the tail stump sperm defect [90]. A splice donor variant in *CCDC189* can cause male sterility in the Nordic Red cattle by disrupting the movement of the sperm flagella [91]. One report of idiopathic male subfertility in Fleckvieh showed that a loss-of-function mutation in *TMEM95* impacts the bull fertility but does not affect semen quality [92].

2 Overview of hemophilia B Leyden

Hemophilia B is a recessive, X-linked bleeding disorder characterized by a deficiency of coagulation factor IX, caused by a different type of mutation in the *F9* gene. Hemophilia B Leyden is a rare and unusual form of hemophilia B in which the causative mutations are all found in the *F9* promoter sequence [93]. In 1970, this specific form of hemophilia B was reported in the Netherlands [94]. In hemophilia B Leyden patients, plasma FIX levels were less than 1% of normal during childhood, but gradually increased after puberty, reaching up to 70% of normal [95]. In humans, mutations identified of hemophilia B Leyden in the *F9* promoter are concentrated in three regions: the first one is about 20 bp upstream of the major transcriptional start site (TSS, +1); the second one is at -5; and the third one is immediately downstream of the TSS, centered around +10 [93]. These three regions correspond to the

binding sites of four transcription factors, namely androgen receptor (AR), hepatocyte nuclear factor 4 α (HNF4 α), one cut homeobox (ONECUT1/2) and CCAAT/enhancing-binding protein α (C/EBP α), respectively [93, 96-98]. These mutations alter the binding site and disrupt the interaction between the transcription factor and the recognition site, thereby affecting the expression of *F9*, at least before puberty. Compared to HNF4 α , AR has been reported to have several alternative binding site motifs [99-101]. As androgen levels rise in childhood, transcript levels of *F9* increase because the intact portion of the promoter is subsequently able to compensate for defective transcription factor binding sites [102, 103].

3 Gene function of Kruppel-like factor 7 (*KLF7*)

Kruppel-like factors (KLFs) are important components of the transcriptional machinery of eukaryotic cells, regulating cell proliferation, differentiation and survival. They are characterized by inclusion of three C₂H₂ zinc fingers at the C-terminus that mediate binding to GC-rich sites. The amino acid sequence of the KLFs' zinc finger region is highly conserved, but the regions beyond the zinc finger are often diverse. KLFs display a wide range of biological properties, each serving as an activator of transcription, a repressor, or both [104, 105]. *KLF7*, also known as the ubiquitous kruppel-like factor (UKLF) due to its structure and ubiquitous expression pattern [106], has been reported to be involved in nervous system and adipogenesis. *KLF7* has been shown to be involved in neuronal morphogenesis and neurogenesis [107, 108]. *KLF7* is also thought to affect neurotransmission, which is an important component of the auditory system [109]. It has been reported that *KLF7* expression can be influenced by the FBXO38 protein [110], and the interaction between FBXO38 and *KLF7* is thought to be associated with the symptoms of hearing loss in human distal hereditary motor neuron disease [111]. *KLF7* has been reported to be a candidate transcription factor of *TCaRE3* (TRKB Ca²⁺ Response Element 3) [112]. TRKB-deficient mice model showed significant hearing loss and TRKB is involved in the remodeling of innervation in the mature cochlea, which means TRKB has potential in the treatment of hearing loss [113]. In addition, *KLF7* has been reported to inhibit preadipocyte differentiation and promotes proliferation of preadipocyte in chickens [114]. One recent report revealed that *KLF7* inhibits chicken adipogenesis through positive regulating of GATA3 expression in preadipocytes [115]. The two opposite directional results imply multiple roles for *KLF7* in transcriptional regulation, consistent with previous findings [104, 105].

Reference

1. Sakaguchi, M., Practical aspects of the fertility of dairy cattle. *Journal of Reproduction and Development*, 2011. 57(1): p. 17-33.
2. Veerkamp, R.F. and B. Beerda, Genetics and genomics to improve fertility in high producing dairy cows. *Theriogenology*, 2007. 68 Suppl 1: p. S266-73.
3. Nebel, R. and M. McGilliard, Interactions of high milk yield and reproductive performance in dairy cows. *Journal of Dairy Science*, 1993. 76(10): p. 3257-3268.
4. Cabrera, V.E., Economics of fertility in high-yielding dairy cows on confined TMR systems. *Animal*, 2014. 8(s1): p. 211-221.
5. Lucy, M., Reproductive loss in high-producing dairy cattle: where will it end? *Journal of dairy science*, 2001. 84(6): p. 1277-1293.
6. Caraviello, D. et al., Survey of management practices on reproductive performance of dairy cattle on large US commercial farms. *Journal of dairy science*, 2006. 89(12): p. 4723-4735.
7. Abdollahi-Arpanahi, R., G. Morota, and F. Penagaricano, Predicting bull fertility using genomic data and biological information. *J Dairy Sci*, 2017. 100(12): p. 9656-9666.
8. Kastelic, J.P., Understanding and evaluating bovine testes. *Theriogenology*, 2014. 81(1): p. 18-23.
9. Parkinson, T.J., Evaluation of fertility and infertility in natural service bulls. *Veterinary Journal*, 2004. 168(3): p. 215-229.
10. Neely, J. et al., Genetic parameters for testes size and sperm number in Hereford bulls. *Journal of Animal Science*, 1982. 55(5): p. 1033-1040.
11. Waldner, C.L., R.I. Kennedy, and C.W. Palmer, A description of the findings from bull breeding soundness evaluations and their association with pregnancy outcomes in a study of western Canadian beef herds. *Theriogenology*, 2010. 74(5): p. 871-883.
12. Rodriguez - Martinez, H., Laboratory semen assessment and prediction of fertility: still utopia? *Reproduction in domestic animals*, 2003. 38(4): p. 312-318.
13. Kjaestad, H., E. Ropstad, and K.A. Berg, Evaluation of spermatological parameters used to predict the fertility of frozen bull semen. *Acta Veterinaria Scandinavica*, 1993. 34(3): p. 299-303.
14. Januskauskas, A., A. Johannisson, and H. Rodriguez-Martinez, Subtle membrane changes in cryopreserved bull semen in relation with sperm viability, chromatin structure, and field fertility. *Theriogenology*, 2003. 60(4): p. 743-758.
15. Zhang, B. et al., Sperm characteristics and zona pellucida binding in relation to field fertility of frozen-thawed semen from dairy AI bulls. *International journal of andrology*, 1998. 21(4): p. 207-216.
16. Kathiravan, P. et al., Objective sperm motion analysis to assess dairy bull fertility using computer - aided system - a review. *Reproduction in Domestic Animals*, 2011. 46(1): p. 165-172.
17. Correa, J., M. Pace, and P. Zavos, Relationships among frozen-thawed sperm characteristics assessed via the routine semen analysis, sperm functional tests and fertility of bulls in an artificial insemination program. *Theriogenology*, 1997. 48(5): p. 721-731.
18. Attia, S., T. Katila, and M. Andersson, The effect of sperm morphology and sire fertility on calving rate of Finnish Ayrshire AI bulls. *Reproduction in domestic animals*, 2016. 51(1): p. 54-58.
19. Nagy, S. et al., Sperm chromatin structure and sperm morphology: Their association with fertility in AI-dairy Ayrshire sires. *Theriogenology*, 2013. 79(8): p. 1153-1161.

20. Ghirardosi, M. et al., Relationship between morphological abnormalities in commercial bull frozen semen doses and conception rate. *Andrologia*, 2018. 50(3): p. e12884.
21. Barth, A.D. et al., Effect of narrow sperm head shape on fertility in cattle. *The Canadian Veterinary Journal*, 1992. 33(1): p. 31.
22. Kastelic, J.P. and J.C. Thundathil, Breeding soundness evaluation and semen analysis for predicting bull fertility. *Reprod Domest Anim*, 2008. 43 Suppl 2: p. 368-73.
23. Chenoweth, P. and F. McPherson, Bull breeding soundness, semen evaluation and cattle productivity. *Animal reproduction science*, 2016. 169: p. 32-36.
24. Braundmeier, A. and D.J. Miller, Invited review: the search is on: finding accurate molecular markers of male fertility. *Journal of dairy science*, 2001. 84(9): p. 1915-1925.
25. Rodriguez-Martinez, H., Sperm function in cattle and pigs: morphological and functional aspects. *ARCHIV FUR TIERZUCHT*, 2001. 44: p. 102-113.
26. Ballachey, B.E., D.P. Evenson, and R.G. SAACKE, The sperm chromatin structure assay relationship with alternate tests of semen quality and heterospermic performance of bulls. *Journal of andrology*, 1988. 9(2): p. 109-115.
27. Thundathil, J. et al., Relationship between the proportion of capacitated spermatozoa present in frozen-thawed bull semen and fertility with artificial insemination. *international journal of andrology*, 1999. 22(6): p. 366-373.
28. Januskauskas, A. et al., Assessment of sperm characteristics post-thaw and response to calcium ionophore in relation to fertility in Swedish dairy AI bulls. *Theriogenology*, 2000. 53(4): p. 859-875.
29. Ward, F. et al., Paternal influence on the time of first embryonic cleavage post insemination and the implications for subsequent bovine embryo development in vitro and fertility in vivo. *Molecular Reproduction and Development: Incorporating Gamete Research*, 2001. 60(1): p. 47-55.
30. Zhang, B. et al., Relationship between embryo development in vitro and 56-day nonreturn rates of cows inseminated with frozen-thawed semen from dairy bulls. *Theriogenology*, 1997. 48(2): p. 221-231.
31. Walters, A. et al., Bovine embryo development after IVF with spermatozoa having abnormal morphology. *Theriogenology*, 2005. 63(7): p. 1925-1937.
32. Utt, M.D., Prediction of bull fertility. *Animal reproduction science*, 2016. 169: p. 37-44.
33. Amann, R. and J. DeJarnette, Impact of genomic selection of AI dairy sires on their likely utilization and methods to estimate fertility: a paradigm shift. *Theriogenology*, 2012. 77(5): p. 795-817.
34. Norman, H., J. Hutchison, and J. Wright, Sire conception rate: New national AI bull fertility evaluation. *AIPL Res. Rep. SCR1 (7-08)*, 2008.
35. Norman, H.D. et al. *A national sire fertility index*. in *Proc. Dairy Cattle Reproductive Council Conf., Omaha, NE. Dairy Cattle Reproductive Council, Hartland, WI*. 2008.
36. Kuhn, M. and J. Hutchison, Prediction of dairy bull fertility from field data: Use of multiple services and identification and utilization of factors affecting bull fertility. *Journal of dairy science*, 2008. 91(6): p. 2481-2492.
37. Nieschlag, E. and H.M. Behre, *Male Reproductive Health and Dysfunction*. 2011, Springer.
38. Agarwal, A. et al., A unique view on male infertility around the globe. *Reproductive biology and endocrinology*, 2015. 13(1): p. 37.
39. Krausz, C., A.R. Escamilla, and C. Chianese, Genetics of male infertility: from research to clinic. *Reproduction*, 2015. 150(5): p. R159-R174.

40. Diskin, M. and D. Morris, Embryonic and early foetal losses in cattle and other ruminants. *Reproduction in Domestic Animals*, 2008. 43: p. 260-267.
41. Krausz, C., Male infertility: pathogenesis and clinical diagnosis. *Best practice & research Clinical endocrinology & metabolism*, 2011. 25(2): p. 271-285.
42. Tournaye, H., C. Krausz, and R.D. Oates, Novel concepts in the aetiology of male reproductive impairment. *The Lancet Diabetes & Endocrinology*, 2017. 5(7): p. 544-553.
43. Krausz, C. and A. Riera-Escamilla, Genetics of male infertility. *Nat Rev Urol*, 2018. 15(6): p. 369-384.
44. Steffen, D., Genetic causes of bull infertility. *Veterinary Clinics: Food Animal Practice*, 1997. 13(2): p. 243-253.
45. Van Camp, S.D., Common causes of infertility in the bull. *Veterinary Clinics: Food Animal Practice*, 1997. 13(2): p. 203-231.
46. Chenoweth, P.J., Genetic sperm defects. *Theriogenology*, 2005. 64(3): p. 457-68.
47. Givens, M.D. and M. Marley, Pathogens that cause infertility of bulls or transmission via semen. *Theriogenology*, 2008. 70(3): p. 504-507.
48. Rolf, M.M. et al., Selection for bull fertility: a review. *Translational Animal Science*, 2020. 4(1): p. 423-441.
49. Fortes, M.R. et al., Candidate genes associated with testicular development, sperm quality, and hormone levels of inhibin, luteinizing hormone, and insulin-like growth factor 1 in Brahman bulls. *Biol Reprod*, 2012. 87(3): p. 58.
50. Hering, D.M., K. Olenski, and S. Kaminski, Genome-wide association study for poor sperm motility in Holstein-Friesian bulls. *Anim Reprod Sci*, 2014. 146(3-4): p. 89-97.
51. Suchocki, T. and J. Szyda, Genome-wide association study for semen production traits in Holstein-Friesian bulls. *Journal of dairy science*, 2015. 98(8): p. 5774-5780.
52. Azcona, F. et al., 155 Whole genome association analysis suggests an influence of inbreeding on bull sperm morphometry. *Reproduction, Fertility and Development*, 2019. 31(1): p. 202-203.
53. Hiltpold, M. et al., Activation of cryptic splicing in bovine WDR19 is associated with reduced semen quality and male fertility. *bioRxiv*, 2020.
54. Feugang, J.M. et al., Two-stage genome-wide association study identifies integrin beta 5 as having potential role in bull fertility. *BMC Genomics*, 2009. 10: p. 176.
55. Whiston, R. et al., A dual targeted β -defensin and exome sequencing approach to identify, validate and functionally characterise genes associated with bull fertility. *Scientific reports*, 2017. 7(1): p. 1-13.
56. Blaschek, M. et al., A whole-genome association analysis of noncompensatory fertility in Holstein bulls. *Journal of dairy science*, 2011. 94(9): p. 4695-4699.
57. Khatib, H. et al., Validation of in vitro fertility genes in a Holstein bull population. *Journal of dairy science*, 2010. 93(5): p. 2244-2249.
58. Peñagaricano, F., K. Weigel, and H. Khatib, Genome - wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. *Animal genetics*, 2012. 43: p. 65-71.
59. Khatib, H., *Single nucleotide polymorphisms associated with bull fertility*. 2015, Google Patents.
60. Han, Y. and F. Peñagaricano, Unravelling the genomic architecture of bull fertility in Holstein cattle. *BMC genetics*, 2016. 17(1): p. 143.
61. Gergely, A. et al., Sperm creatine kinase activity in normospermic and oligospermic Hungarian men. *Journal of assisted reproduction and genetics*, 1999. 16(1): p. 35-40.
62. Shoji, M. et al., The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Developmental cell*, 2009. 17(6): p. 775-787.

63. Griffeth, R.J., V. Bianda, and S. Nef, The emerging role of insulin-like growth factors in testis development and function. *Basic and clinical andrology*, 2014. 24(1): p. 12.
64. Thomas, T., K.L. Loveland, and A.K. Voss, The genes coding for the MYST family histone acetyltransferases, Tip60 and Mof, are expressed at high levels during sperm development. *Gene expression patterns*, 2007. 7(6): p. 657-665.
65. Dun, M.D. et al., The chaperonin containing TCP1 complex (CCT/TRiC) is involved in mediating sperm-oocyte interaction. *Journal of Biological Chemistry*, 2011. 286(42): p. 36875-36887.
66. Nicolini, P. et al., Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. *BMC genetics*, 2018. 19(1): p. 14.
67. Wolfsberg, T.G. et al., ADAM, a Widely Distributed and Developmentally Regulated Gene Family Encoding Membrane Proteins with A Δ Disintegrin A Δ Metalloprotease Domain. *Developmental biology*, 1995. 169(1): p. 378-383.
68. Oh, J. et al., Molecular, biochemical, and cellular characterization of epididymal ADAMs, ADAM7 and ADAM28. *Biochemical and biophysical research communications*, 2005. 331(4): p. 1374-1383.
69. Terada, K. et al., A type I DnaJ homolog, DjA1, regulates androgen receptor signaling and spermatogenesis. *The EMBO journal*, 2005. 24(3): p. 611-622.
70. Liegel, R.P. et al., Loss-of-function mutations in TBC1D20 cause cataracts and male infertility in blind sterile mice and Warburg micro syndrome in humans. *The American Journal of Human Genetics*, 2013. 93(6): p. 1001-1014.
71. Keeney, S., C.N. Giroux, and N. Kleckner, Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, 1997. 88(3): p. 375-384.
72. Paronetto, M.P. and C. Sette, Role of RNA - binding proteins in mammalian spermatogenesis. *International journal of andrology*, 2010. 33(1): p. 2-12.
73. Mizuhashi, K. et al., Obif, a transmembrane protein, is required for bone mineralization and spermatogenesis in mice. *PloS one*, 2015. 10(7).
74. Rezende, F., G. Dietsch, and F. Peñagaricano, Genetic dissection of bull fertility in US Jersey dairy cattle. *Animal genetics*, 2018. 49(5): p. 393-402.
75. Nani, J.P. and F. Penagaricano, Whole-genome homozygosity mapping reveals candidate regions affecting bull fertility in US Holstein cattle. *BMC Genomics*, 2020. 21(1): p. 338.
76. Krawetz, S.A., Paternal contribution: new insights and future challenges. *Nature Reviews Genetics*, 2005. 6(8): p. 633-642.
77. Kropp, J. et al., Male fertility status is associated with DNA methylation signatures in sperm and transcriptomic profiles of bovine preimplantation embryos. *BMC genomics*, 2017. 18(1): p. 280.
78. Lalancette, C. et al., Transcriptome analysis of bull semen with extreme nonreturn rate: use of suppression-subtractive hybridization to identify functional markers for fertility. *Biology of reproduction*, 2008. 78(4): p. 618-635.
79. Arangasamy, A. et al., Association of CRISP2, CCT8, PEBP1 mRNA abundance in sperm and sire conception rate in Holstein bulls. *Theriogenology*, 2011. 76(3): p. 570-577.
80. Card, C.J. et al., Oligo-dT selected spermatozoal transcript profiles differ among higher and lower fertility dairy sires. *Animal Reproduction Science*, 2017. 177: p. 105-123.
81. Govindaraju, A. et al., Dynamics of microRNAs in bull spermatozoa. *Reproductive Biology and Endocrinology*, 2012. 10(1): p. 82.

82. Gross, N., F. Peñagaricano, and H. Khatib, Integration of whole - genome DNA methylation data with RNA sequencing data to identify markers for bull fertility. *Animal genetics*, 2020. 51(4): p. 502-510.
83. Peddinti, D. et al., Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. *BMC systems biology*, 2008. 2(1): p. 19.
84. D'Amours, O. et al., Proteomic comparison of detergent-extracted sperm proteins from bulls with different fertility indexes. *Reproduction*, 2010. 139(3): p. 545-556.
85. Bellin, M.E., H.E. Hawkins, and R.L. Ax, Fertility of range beef bulls grouped according to presence or absence of heparin-binding proteins in sperm membranes and seminal fluid. *Journal of animal science*, 1994. 72(9): p. 2441-2448.
86. Sprott, L. et al., Artificial insemination outcomes in beef females using bovine sperm with a detectable fertility-associated antigen. *Journal of animal science*, 2000. 78(4): p. 795-798.
87. Parent, S. et al., Bull subfertility is associated with low levels of a sperm membrane antigen. *Molecular reproduction and development*, 1999. 52(1): p. 57-65.
88. Dai, L. et al., Effects of novel single nucleotide polymorphisms of the FSH beta-subunit gene on semen quality and fertility in bulls. *Anim Reprod Sci*, 2009. 114(1-3): p. 14-22.
89. Dai, L. et al., Naturally occurring genetic mutations in the 5'-upstream regulatory region of bovine FSHB generate a novel cis-regulatory element that affects its expression. *Anim Genet*, 2015. 46(6): p. 693-6.
90. Pausch, H. et al., A frameshift mutation in ARMC3 is associated with a tail stump sperm defect in Swedish Red (*Bos taurus*) cattle. *BMC Genet*, 2016. 17: p. 49.
91. Iso-Touru, T. et al., A splice donor variant in CCDC189 is associated with asthenospermia in Nordic Red dairy cattle. *BMC Genomics*, 2019. 20(1): p. 286.
92. Pausch, H. et al., A nonsense mutation in TMEM95 encoding a nondescript transmembrane protein causes idiopathic male subfertility in cattle. *PLoS Genet*, 2014. 10(1): p. e1004044.
93. Funnell, A.P. and M. Crossley, Hemophilia B Leyden and once mysterious cis-regulatory mutations. *Trends in Genetics*, 2014. 30(1): p. 18-23.
94. Veltkamp, J. et al., Another genetic variant of haemophilia B: haemophilia B Leyden. *Scandinavian journal of haematology*, 1970. 7(2): p. 82-90.
95. Giannelli, F. et al., Haemophilia B: database of point mutations and short additions and deletions—third edition, 1992. *Nucleic acids research*, 1992. 20(Suppl): p. 2027.
96. Funnell, A.P. et al., A CpG mutational hotspot in a ONECUT binding site accounts for the prevalent variant of hemophilia B Leyden. *The American Journal of Human Genetics*, 2013. 92(3): p. 460-467.
97. Portales-Casamar, E. et al., JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic acids research*, 2010. 38(suppl_1): p. D105-D110.
98. Tewari, A.K. et al., Chromatin accessibility reveals insights into androgen receptor activation and transcriptional specificity. *Genome biology*, 2012. 13(10): p. 1-17.
99. Reid, K.J. et al., Two classes of androgen receptor elements mediate cooperativity through allosteric interactions. *Journal of Biological Chemistry*, 2001. 276(4): p. 2943-2952.
100. Schuur, E.R. et al., Prostate-specific antigen expression is regulated by an upstream enhancer. *Journal of Biological Chemistry*, 1996. 271(12): p. 7043-7051.
101. Wilson, S., J. Qi, and F.V. Filipp, Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines. *Scientific reports*, 2016. 6(1): p. 1-15.

102. Rallapalli, P. et al., An interactive mutation database for human coagulation factor IX provides novel insights into the phenotypes and genetics of hemophilia B. *Journal of thrombosis and haemostasis*, 2013. 11(7): p. 1329-1340.
103. Reijnen, M.J. et al., Hemophilia B Leyden: substitution of thymine for guanine at position-21 results in a disruption of a hepatocyte nuclear factor 4 binding site in the factor IX promoter. 1993.
104. Dang, D.T., J. Pevsner, and V.W. Yang, The biology of the mammalian Krüppel-like family of transcription factors. *The international journal of biochemistry & cell biology*, 2000. 32(11-12): p. 1103-1121.
105. Kaczynski, J., T. Cook, and R. Urrutia, Sp1-and Krüppel-like transcription factors. *Genome biology*, 2003. 4(2): p. 1-8.
106. Matsumoto, N. et al., Cloning the cDNA for a new human zinc finger protein defines a group of closely related Krüppel-like transcription factors. *Journal of Biological Chemistry*, 1998. 273(43): p. 28229-28237.
107. Laub, F. et al., Transcription factor KLF7 is important for neuronal morphogenesis in selected regions of the nervous system. *Molecular and Cellular Biology*, 2005. 25(13): p. 5699-5711.
108. Laub, F. et al., Developmental expression of mouse Krüppel-like transcription factor KLF7 suggests a potential role in neurogenesis. *Developmental biology*, 2001. 233(2): p. 305-318.
109. Kajimura, D. et al., Identification of genes regulated by transcription factor KLF7 in differentiating olfactory sensory neurons. *Gene*, 2007. 388(1-2): p. 34-42.
110. Sumner, C.J. et al., A dominant mutation in FBXO38 causes distal spinal muscular atrophy with calf predominance. *The American Journal of Human Genetics*, 2013. 93(5): p. 976-983.
111. Akçimen, F. et al., A novel homozygous FBXO38 variant causes an early-onset distal hereditary motor neuropathy type IID. *Journal of human genetics*, 2019. 64(11): p. 1141-1144.
112. Kingsbury, T.J. and B.K. Krueger, Ca²⁺, CREB and krüppel: a novel KLF7-binding element conserved in mouse and human TRKB promoters is required for CREB-dependent transcription. *Molecular and Cellular Neuroscience*, 2007. 35(3): p. 447-455.
113. Schimmang, T. et al., Lack of Bdnf and TrkB signalling in the postnatal cochlea leads to a spatial reshaping of innervation along the tonotopic axis and hearing loss. *Development*, 2003. 130(19): p. 4741-4750.
114. Zhang, Z. et al., Klf7 modulates the differentiation and proliferation of chicken preadipocyte. *Acta Biochim Biophys Sin*, 2013. 45(4): p. 280-288.
115. Sun, Y. et al., GATA binding protein 3 is a direct target of Kruppel-like transcription factor 7 and inhibits chicken adipogenesis. *Frontiers in Physiology*, 2020. 11: p. 610.

CHAPTER 2

Association of α/β -Hydrolase D16B with Bovine Conception Rate and Sperm Plasma Membrane Lipid Composition

Shuwen Shan[†], Fangzheng Xu[†], Martina Bleyer, Svenja Becker, Torben Melbaum, Wilhelm Wemheuer, Marc Hirschfeld, Christin Wacker, Shuhong Zhao, Ekkehard Schütz, and Bertram Brenig^{*}

[†] These authors contributed equally to this work.

^{*} Corresponding author

The article was published online in International journal of molecular sciences, 21(2), p.627. in January 2020. The full article can be found online at:
<https://doi.org/10.3390/ijms21020627>

Author contributions:

Shuwen Shan participated in the GWAS and WGS data analysis, gene conservation analysis and western blot. She also involved in the manuscript writing and project discussion.

Acknowledgments:

Shuwen Shan would like to thank Zentrum für molekulare Diagnostik (ZMD) for generating genomic sequencing data of the project.



Article

Association of α/β -Hydrolase D16B with Bovine Conception Rate and Sperm Plasma Membrane Lipid Composition

Shuwen Shan ^{1,†}, Fangzheng Xu ^{1,†}, Martina Bleyer ², Svenja Becker ¹, Torben Melbaum ¹, Wilhelm Wemheuer ¹, Marc Hirschfeld ^{1,3} , Christin Wacker ¹, Shuhong Zhao ⁴, Ekkehard Schütz ¹ and Bertram Brenig ^{1,*}

¹ Institute of Veterinary Medicine, University of Goettingen, 37077 Goettingen, Germany; shuwen.shan@stud.uni-goettingen.de (S.S.); fangzheng.xu@stud.uni-goettingen.de (F.X.); svenja_b_@web.de (S.B.); torbenmelbaum@web.de (T.M.); wwemheu1@gwdg.de (W.W.); marc.hirschfeld@uni-goettingen.de (M.H.); cwehrha@gwdg.de (C.W.); eschuet@gwdg.de (E.S.)

² Pathology Unit, German Primate Center, Leibniz-Institute for Primate Research Goettingen, 37077 Goettingen, Germany; MBleyer@dpz.eu

³ Department of Obstetrics and Gynecology, University Medical Center Freiburg, 79106 Freiburg, Germany

⁴ Key Lab of Animal Genetics, Breeding and Reproduction, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, China; shzhao@mail.hzau.edu.cn

* Correspondence: bbrenig@gwdg.de; Tel.: +49-551-3928383; Fax: +49-551-3933392

† These authors contributed equally to this work.

Received: 3 January 2020; Accepted: 15 January 2020; Published: 17 January 2020



Abstract: We have identified a Holstein sire named Tarantino who had been approved for artificial insemination that is based on normal semen characteristics (i.e., morphology, thermoresistance, motility, sperm concentration), but had no progeny after 412 first inseminations, resulting in a non-return rate (NR_{dev}) of -29 . Using whole genome association analysis and next generation sequencing, an associated nonsense variant in the α/β -hydrolase domain-containing 16B gene (*ABHD16B*) on bovine chromosome 13 was identified. The frequency of the mutant allele in the German Holstein population was determined to be 0.0018 in 222,645 investigated cattle specimens. The mutant allele was traced back to Whirlhill Kingpin (born Feb. 13th, 1959) as potential founder. The expression of *ABHD16B* was detected by Western blotting and immunohistochemistry in testis and epididymis of control bulls. A lipidome comparison of the plasma membrane of fresh semen from carriers and controls showed significant differences in the concentration of phosphatidylcholine (PC), diacylglycerol (DAG), ceramide (Cer), sphingomyelin (SM), and phosphatidylcholine (-ether) (PC O-), indicating that *ABHD16B* plays a role in lipid biosynthesis. The altered lipid contents may explain the reduced fertilization ability of mutated sperms.

Keywords: Holstein cattle; male infertility; *ABHD16B*

1. Introduction

Fertility is an important economical productivity factor in animal breeding [1–4]. Indicators to assess male fertility can either be indirect (e.g., productivity of progeny, sire conception rate, non-return rate) or direct (e.g., semen characteristics, testis size) [5–10]. The latter parameters have the advantage that they can be easily measured and they provide an immediate answer; however, the heritabilities of scrotal circumference and semen traits vary extremely, ranging from 0.0 (i.e., abnormal heads, bent tails, distal cytoplasmic droplets) to 0.57 (i.e., scrotal circumference) and, therefore, their use in selection is not always straightforward [11]. Alternative approaches were used to determine the differences

between fertile and infertile bulls while using molecular tools. Transcriptome analyses for instance have shown that spermatozoa of high-fertility bulls show a higher concentration of specific transcripts for membrane and extracellular space protein locations [12,13]. In another study residual RNA content in spermatozoa of bulls with extreme non-return rates was analysed [14]. Low-fertile bulls showed a significantly increased amount of ribosomal and mitochondrial sequences, whereas high-fertile bulls exhibited transcripts of genes that are involved, for example, in metabolism, signal transduction, translation, and protein degradation [14]. From transcriptome and proteome studies, mainly in man, mouse, and rat, it is evident that differences between RNA and protein content, DNA methylation, posttranslational modifications between fertile and infertile individuals exist [15–18]. The use of these types of biomarkers in reproductive medicine is believed to bridge the gap between conventional semen analysis with limited clinical utility and biochemical pathways that regulate male fertility [19].

However, the assessment of mutational effects in candidate genes is normally challenging, especially when there are only subtle deviations in expression levels, due to the complex interactions of geno- and phenotypes in fertility traits [20,21]. With the advancement of high-throughput screening tools (DNA chip, next generation sequencing) and the availability of large datasets on fertility parameters of bulls, especially in Holstein cattle male fertility, can be practically implemented into genomic selection [22]. Genome-wide association studies have been conducted in Holstein bulls, identifying several fertility associated genomic regions [23]. A recent genome-wide association study has detected at least eight genomic regions, i.e., on bovine chromosome 5 (BTA5), BTA9, BTA13, BTA21, and BTA25, in Holstein cattle associated with bull fertility while using Sire Conception Rate (SCR) as a parameter [24]. In a large multi-species comparative study 33 promising candidate genes have been identified for male fertility/infertility [25]. Recently, a whole exome sequencing of 24 high and low fertile bulls identified 484 SNPs that were significantly associated with fertility [26]. The second most significantly associated SNP in this study was located on BTA13 at position 53,691,419 within the *SIRPA* gene. Although these data point at a number of potential molecular targets only three causative mutations, resulting in male sub- or infertility in cattle have been determined in the *FSHB*, *TMEM95*, and *ARMC3* gene hitherto [27–30].

Here, we report about the identification of nonsense variant in the bovine α/β -hydrolase D16B gene (*ABHD16B*) on BTA13 significantly associated with male subfertility in Holstein cattle. So far, nothing was known regarding the physiological or biochemical function of *ABHD16B* [31,32]. Our data provide evidence that *ABHD16B* is involved in lipid biosynthesis in testis and is crucial for fertilization.

2. Results

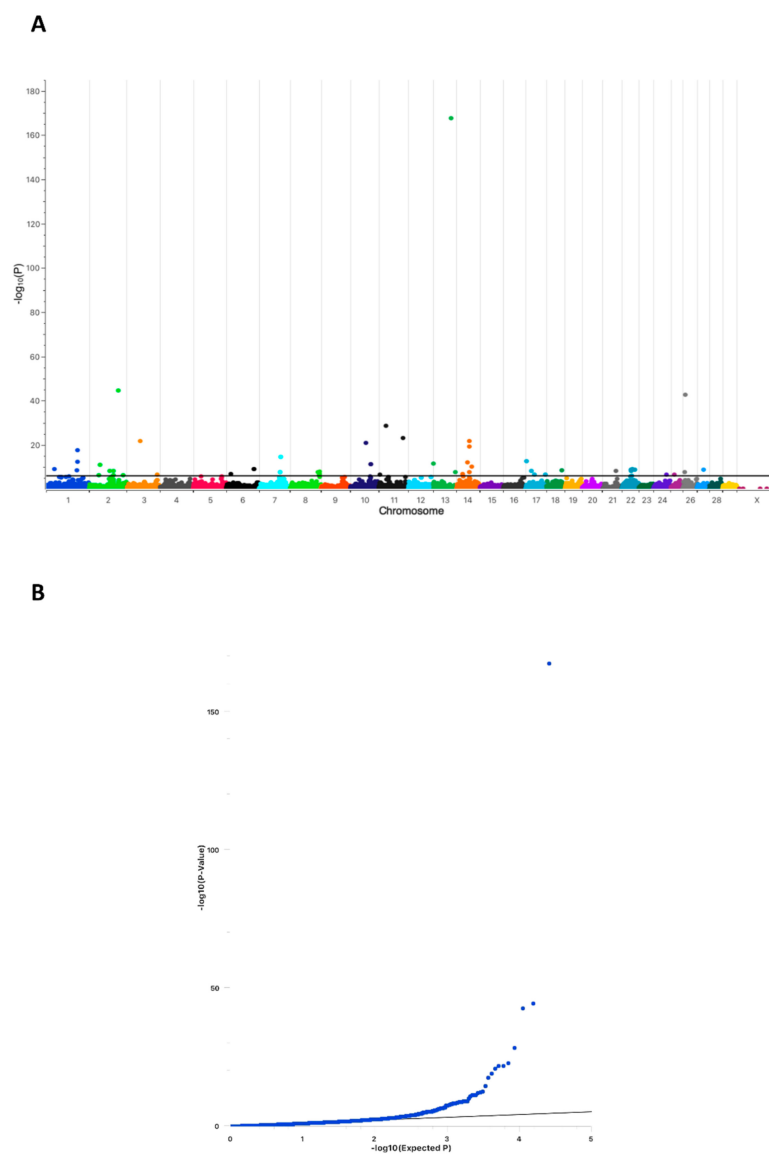
2.1. Conception Ability of Sires Is Highly Associated with a Chromosomal Region on Bovine Chromosome 13

A Genome Wide Association Analysis (GWAS) was performed while using a cohort of 289 Holstein sires to determine chromosomal regions harboring associated causative genes for conception ability (NR_{dev}). The cohort consisted of 10 sires with a $NR_{dev} \leq -2$ (= cases) (Table 1) and 279 randomly chosen sires of the active breeding population (= controls).

Individual NR_{dev} values of the control sires were not available; however, they were assumed to be normal, as all of these sires were used in the current breeding population. As shown in Figure 1A, one genome-wide highly significant associated position on BTA13 (ARS-BFGL-NGS-107931; position 63,500,701) was detected ($-\log_{10}P$ -value = 167.56). Seventeen additional regions above a Bonferroni threshold of $-\log_{10}P = 5.9$ ($p < 0.01$) with much lower significance were present on BTA1, 2, 3, 6, 7, 8, 10, 11, 14, 17, 18, 21, 22, 24, 25, 26, and 27. The QQ-plot clearly indicated a compelling evidence for an excess of association with no population substructure (Figure 1B). Whole-genome sequencing was performed while using Tarantino and his parents to determine which of the associated chromosomal regions harbored protein-altering variants that were causative for Tarantino's infertility.

Table 1. Sub- and infertile sires selected for genome-wide association analysis.

Sire/ID	NR _{dev} ^{a)}	No. of First Inseminations
Tarantino	−29	412
19_39644	−27	402
19_39643	−25	364
05_34345	−9	412
04_44565	−4	421
04_39067	−3	315
04_43327	−3	424
04_40476	−2	407
04_41962	−2	640
04_37666	−2	571

a) NR_{dev}: Non-return rate deviation.**Figure 1.** Manhattan plot of the Genome Wide Association Analysis (GWAS) ($n = 289$; 279 controls, 10 cases). (A) The plot shows the $-\log_{10}$ -transformed p -values for all SNPs. The black horizontal line represents the genome-wide significance threshold of $-\log_{10}P = 5.9$. (B) Quantile-quantile (QQ) plot of the GWAS.

2.2. Whole-Genome Sequencing Reveals Two Potential Protein-Altering Variants Upstream the Associated Position on BTA13

Raw next generation sequencing data were quality filtered. Within the filtered 78,472 SNPs, only 20 resulted in a predicted loss of function, including 10 nonsense variants, five splice-donor variants, three splice acceptor-variants, and two initiator-codon variants. Two SNPs were located near the associated position on BTA13, i.e., a nonsense variant at position 54,429,815 within the single exonic α/β -hydrolase D16B (*ABHD16B*) gene (AC_000170.1: g.54429815G>A, rs468948776) and a splice-acceptor variant at position 53,003,648 within the transmembrane channel-like protein 2 (*TMC2*) gene (AC_000170.1: g.53003648C>T, rs465702794). *TMC2* has been shown to be expressed in the inner ear and it is necessary for the mechanotransduction in cochlear hair cells [33,34]. *TMC2* was excluded as potential candidate due to this very specific function. On the other hand, *ABHD16B* has been shown in humans to be mainly expressed in testis, which suggested a potential role in Tarantino's infertility [35]. In addition, aberrant methylation patterns of *ABHD16B* have been shown to be associated with infertility in men [36].

2.3. Verification and Validation of the Nonsense Variant g.54429815G>A (*ABHD16B*) in the Holstein Population

An initial set of 2072 randomly selected Holstein DNA samples were genotyped to verify and validate the presence of the detected variant in *ABHD16B*. In this set, 2052 wild type (G_G), 20 heterozygous (G_A), and no homozygous (A_A) carrier were detected (HWE $\chi^2 = 0.05$). The results proved that the variant was present in the population at a very low frequency. Therefore, a larger cohort of 222,645 HF cattle (208,165 female, 14,480 male) was genotyped while using the bovinSNP50 BeadChip. In this cohort, 810 heterozygous (781 female, 19 male, 10 unknown sex) and no homozygous animals were identified, resulting in a frequency of the variant allele of 0.0018. According to Hardy–Weinberg equilibrium it was not unexpected that no homozygous individuals were detected (HWE $\chi^2 = 0.73$). The low allele frequency further supported the data that the nonsense variant in *ABHD16B* was most likely the causative variant for Tarantino's infertility, because sires will be rapidly removed from the breeding population once a sub- or infertility would have been evident during routine fertility testing. Such a selection will efficiently reduce the transmission and spreading of the causative variant. The limited number of heterozygous individuals in the randomly chosen large Holstein cohort prompted us to determine the number of heterozygous sires in the available DNA samples of Tarantino's close male relatives in correlation with their conception ability (NR_{dev}). A total of 34 DNA samples were available and genotyped, resulting in 16 wild type and 18 heterozygous sires (HWE $\chi^2 = 4.4$). Within the heterozygous sires, 15 had negative NR_{dev} values (-9 to < 0) and only three sires showed positive NR_{dev} values (0 to 2).

2.4. Expression and Tissue Distribution of *ABHD16B*

ABHD16B codes for a protein of 470 amino acids with a predicted α/β -hydrolase fold domain. The nonsense variant g.54429815G>A causes a premature stop at amino acid position 218 (glutamine residue), resulting in a truncation of 253 C-terminal amino acids and 53.8% of the protein (Figure 2). In silico protein sequence comparison of 11 mammals revealed that the glutamine residue (Q) is highly conserved. Due to the truncation 67.4% of the α/β -hydrolase fold domain is missing. Regarding the evolutionary appearance, it is interesting to note that *ABHD16B* first evolved in reptiles performing internal fertilization. Species with external fertilization, e.g., fish and frogs, do not harbor an *ABHD16B* gene.

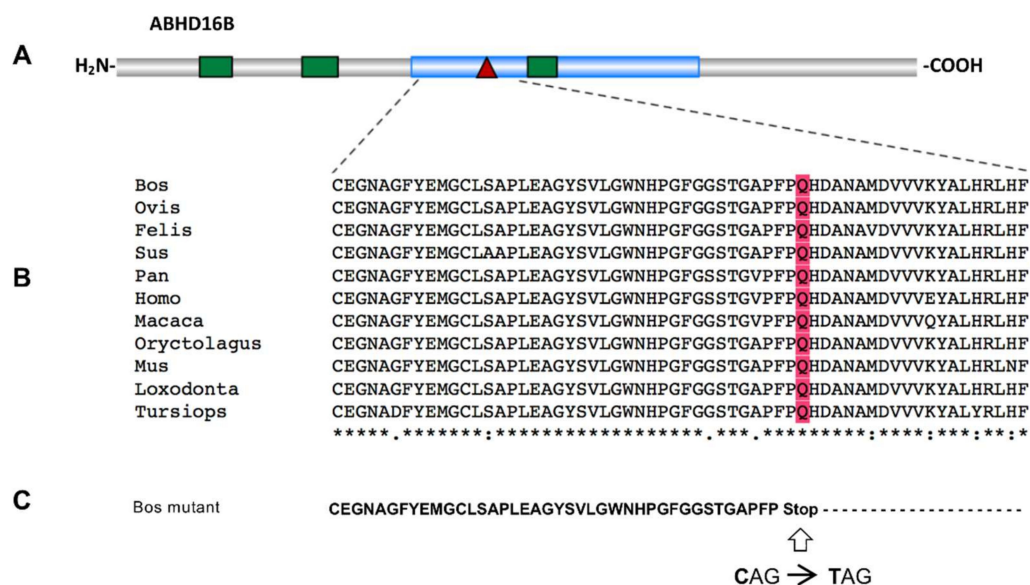


Figure 2. (A) Schematic representation of the ABHD16B protein structure indicates the position of the α/β -hydrolase fold domain (blue) and the transmembrane helices (green), predicted by NCBI Conserved Domains Database and TMbase, respectively. The amino acid position (218) of the nonsense variant leading in a premature stop is marked by a red triangle. **(B)** The comparative alignment of amino acid sequences of 11 mammals while using Clustal W (178-amino acid position 178–237) is shown. The amino acid position at the truncation site is indicated in red. NCBI protein sequence accession numbers are as follows: Bos (*Bos Taurus*) NP_001033630.1, Ovis (*Ovis aries*) XP_014955258.1, Felis (*Felis catus*) XP_003983341.3, Sus (*Sus scrofa*) XP_020933693.1, Pan (*Pan troglodytes*) XP_003317106.1, Homo (*Homo sapiens*) NP_542189.1, Macaca (*Macaca mulatta*) NP_001180656.1, Oryctolagus (*Oryctolagus cuniculus*) XP_008250767.2, Mus (*Mus musculus*) NP_899004.1, Loxodonta (*Loxodonta Africana*) XP_003421827.1, Tursiops (*Tursiops truncatus*) XP_019806804.1. **(C)** The amino acid sequence of bovine ABHD16B truncated protein with the stop-gain variant.

2.5. ABHD16B Is Expressed in Testis but not in Spermatozoa

Western blotting was used to detect ABHD16B in testis (wild type) and spermatozoa (wild type, heterozygous, and homozygous variant) extracts. Testes of heterozygous and/or homozygous carriers were unavailable due to the low genotype frequencies. However, a limited amount of deep-frozen semen samples of Tarantino and a further not directly related homozygous carrier (C_a) provided from the safety inventory of an AI station were included in the analysis. An ABHD16B specific band was detected in testis of wild type bulls at the expected size of approx. 70 kDa, as shown in Figure 3. Neither in wild type nor in heterozygous or homozygous variant spermatozoa extracts ABHD16B was detected (Figure 3B).

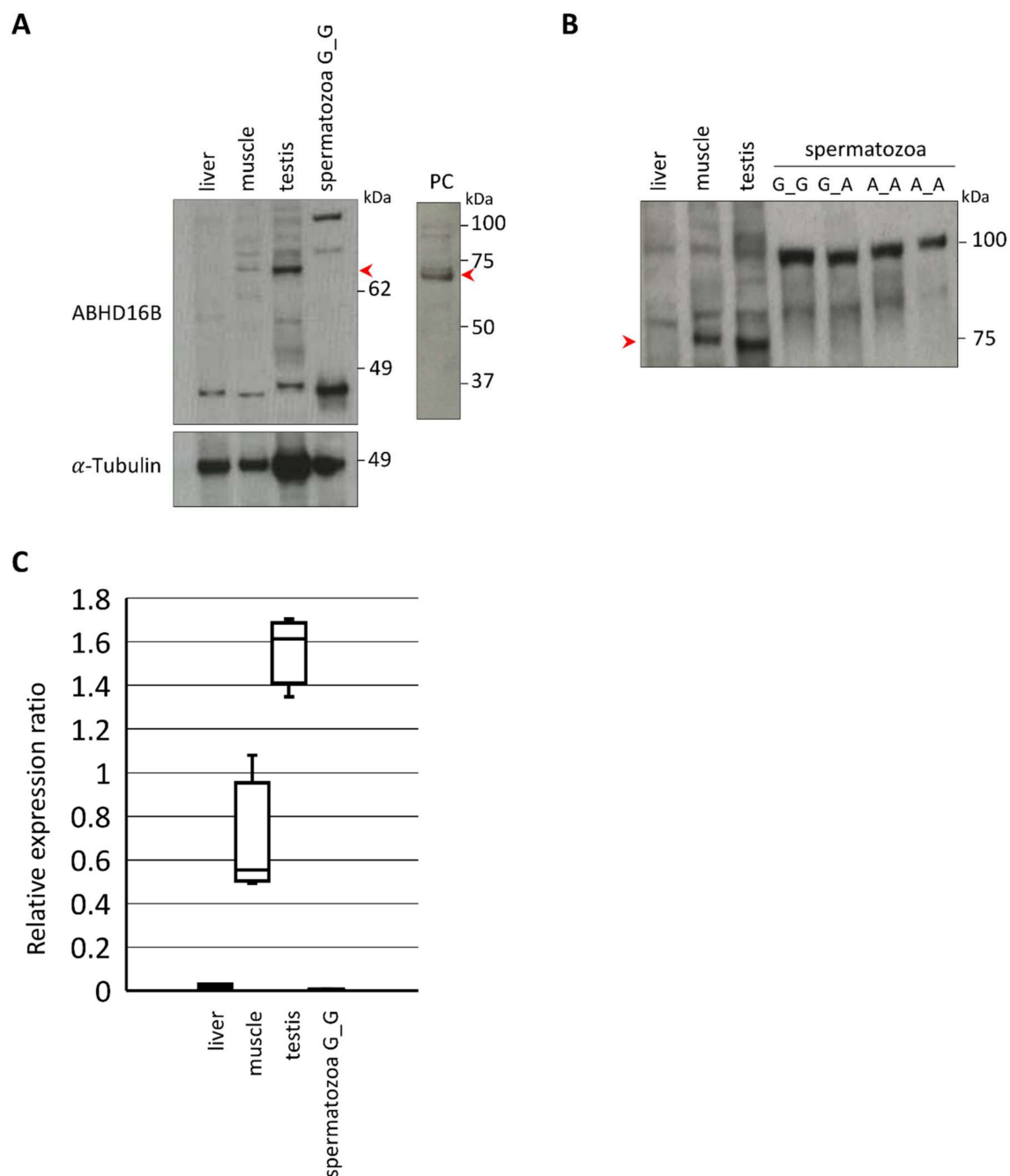


Figure 3. Western blot analysis of ABHD16B protein expression in tissues and spermatozoa. **(A)** ABHD16B protein band (red arrow) detected in wild type testis (G_G) but not in spermatozoa. Using liver as negative control, muscle and human ABHD16B over-expression lysate as positive controls (PC). α-Tubulin used as the loading control. **(B)** ABHD16B (approx. 70 kDa) is absent in spermatozoa of three genotypes (wild type (G_G), heterozygous (G_A) and homozygous carrier (A_A)). Spermatozoa of homozygous carriers were from Tarantino and C_a. **(C)** Box and Whisker plot of relative ABHD16B expression. Areas under curve were determined using ImageJ 1.52k software and relative expression ratios of ABHD16B (%) in liver, muscle, testis, and spermatozoa (G_G) were calculated while using α-Tubulin expression as internal standard. Horizontal lines within boxes indicate median values and whiskers show upper and lower extremes.

2.6. Immunohistochemical Analysis Revealed ABHD16B Expression in Testis and Epididymis

Sections of testicular and epididymal tissue samples that were collected at an abattoir were prepared for IHC. The *ABHD16B* genotype of the samples was tested prior to IHC and shown to originate from wild type sires. While using the PAC-ARK antibody, ABHD16B expression was detectable in testicular parenchyma, ductuli efferentes, as well as epididymal tail, body, and head, as shown in Figure 4. Specifically, there is ABHD16B expression in the nucleoplasm of Leydig cells, in the seminiferous tubules and, with variable intensity, in the epithelium of the ductus epididymis. These findings suggest that ABHD16B probably plays a role in spermatogenesis and sperm maturation.

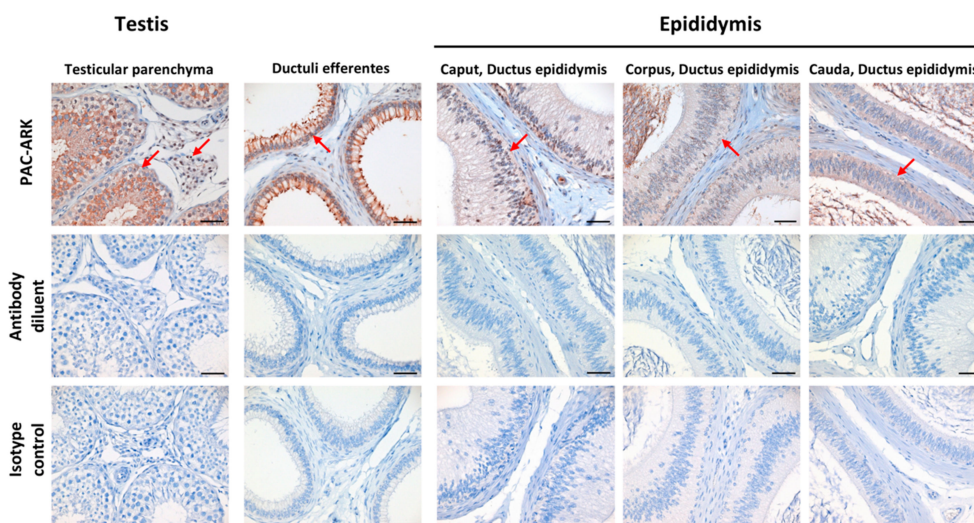


Figure 4. ABHD16B protein detection and localization in testis and epididymis of wild type bull by immunohistochemistry. PAC-ARK antibody was used as primary antibody. Positive staining is indicated with red arrows. When the primary antibody (PAC-ARK) was replaced with antibody diluent and isotype rabbit IgG at the same working dilution, no staining was observed in any of these tissues. Scale bars = 50 μ m.

2.7. ABHD16B Is Involved in Lipid Metabolism and Influences Sperm Plasma Membrane Lipid Composition

We hypothesized that ABHD16B could be involved in plasma membrane lipid biosynthesis, as many members of the α/β -hydrolase superfamily of hydrolytic enzymes are involved in lipid metabolism. Sperm lipidomics of heterozygous and wild type semen samples was performed to interrogate this hypothesis. The number of available semen samples of Tarantino was limited and, therefore, it was decided not to include these valuable samples. However, if ABHD16B would have an effect on lipid biosynthesis during spermatogenesis, this should also be detectable in heterozygous samples. After normalization to 10^6 sperms per sample, no significant difference in the total lipid content between 15 wild type and 15 heterozygous samples was detected (Figure 5).

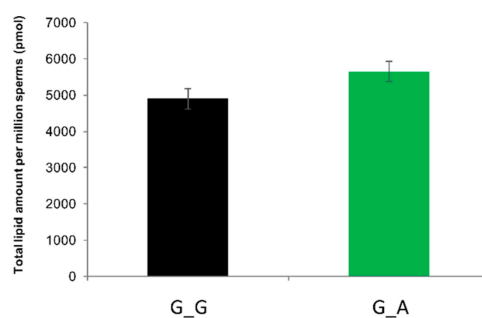


Figure 5. Total lipid amount of 10^6 sperms of each genotype revealed no significant difference after normalization. G_G: wild type; G_A: heterozygous carrier.

However, 10 out of 16 lipid classes showed significant differences (Figure 6A). The majority of different lipids belonged to the classes of diacylglycerols (DAG), glycerophosphocholines (PC, PC O-), ceramides (Cer), and sphingomyelins (SM). The sperms of heterozygous carriers showed significantly decreased amounts of SM and DAG, while PC, PC O-, and Cer were increased (Figure 6A). In total, 99 of 144 lipid species demonstrated significant differences between wild type and heterozygous sperm samples. Eight lipid species significantly decreased ($p_{(BH)} < 0.05$, $\log_2 fc < -1$), six of them were DAGs. 25 lipid species significantly increased ($p_{(BH)} < 0.05$, $\log_2 fc > 1$), almost half ($n=12$) of them were PCs (Figure 6B). Figure 6C shows the ten most significantly changed lipid species. An important indicator of cell membrane integrity is the PC:PE ratio. As shown in Figure 6D, heterozygous sperms have a significantly increased the PC:PE ratio. Another sperm membrane structure criterion is the ratio between LPC 22:6 and PC 16:0_22:6, and it also significantly increased in the heterozygous samples (Figure 6E).

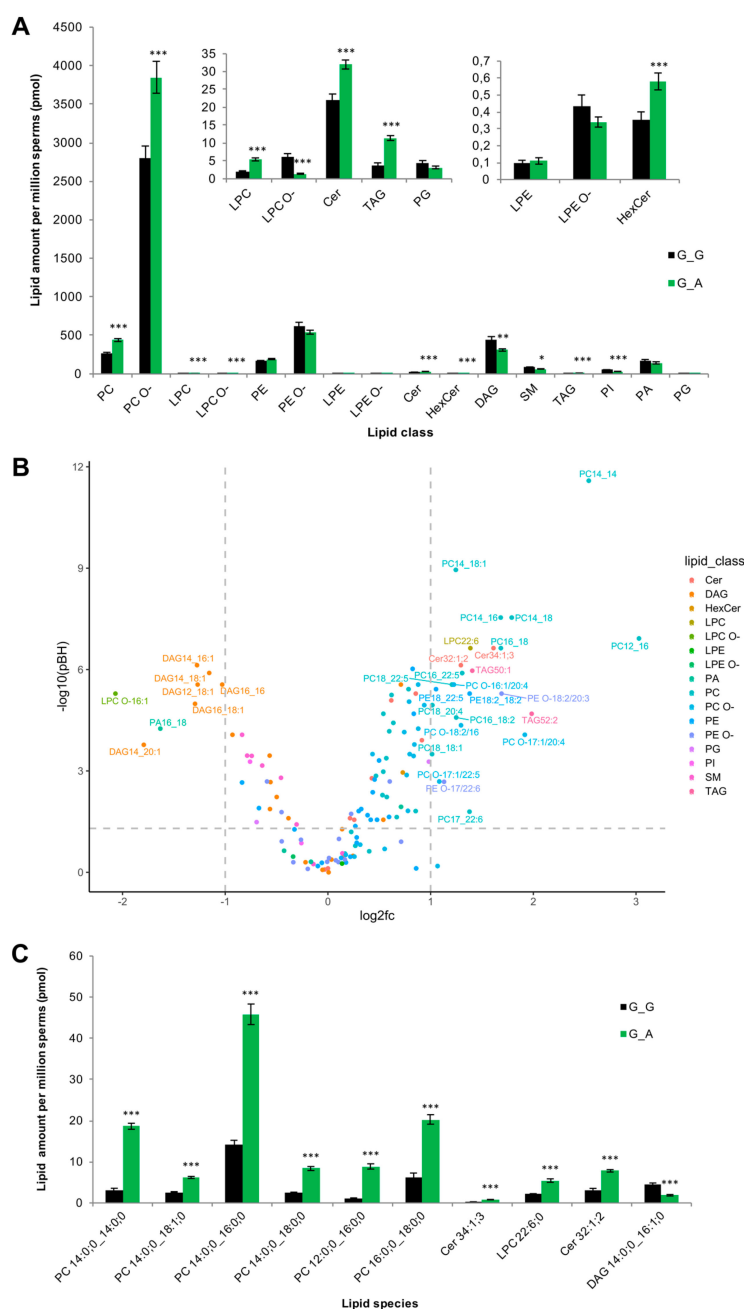


Figure 6. Cont.

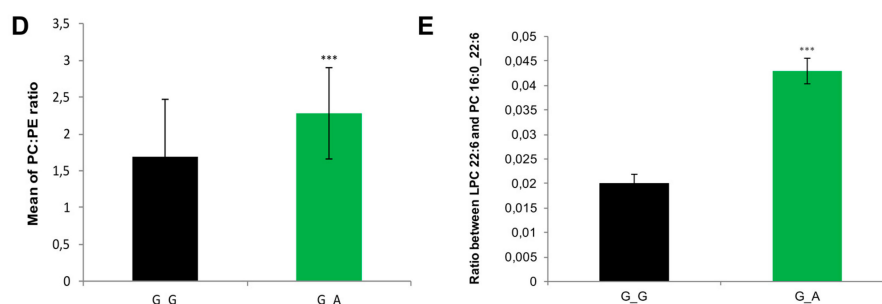


Figure 6. Lipid distribution and variance of spermatozoa in wild type and heterozygous sires. (A) Change profile of total lipid content for each lipid class of sperms in two genotypes. (B) $-\log_{10}$ of adjusted p -value ($p_{(BH)}$) and mean \log_2 -fold change (G_A vs G_G) of 144 lipid species were plotted, lipid species with $-\log_{10}P_{(BH)} > 1.3$ and $|\log_2\text{fold change}| > 1$ (dashed grey lines are included) were annotated with lipid feature names (simplified without the saturated acyl groups). (C) Top 10 most significantly changed lipid species between G_G and G_A bull sperm cells. (D) Bar chart shows the mean of PC:PE ratio in G_G and G_A bull sperms. (E) Bar chart shows the ratio between LPC 22:6 and PC 16:0_22:6 in two genotypes. G_G: wild type; G_A: heterozygous. Differential changes were tested by the Mann–Whitney U test or Student’s t -test. Data are presented as mean \pm SEM; $n = 15/\text{group}$; FDR adjusted p -values are indicated: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3. Discussion

Male infertility is a complex multifactorial idiopathic, congenital, or acquired heterogeneous disease [37,38]. In men, genetic factors predominantly cause idiopathic conditions contributing to 30–40% of male infertility [39]. However, up to now, for only three genes, i.e., *NR5A1*, *DMRT1*, and *TEX11*, associations with male infertility have been evidenced in independent biological and functional studies [40]. The same number of genes has been identified in cattle causing bull sub- or infertility so far, i.e., *FSHB* (BTA15, 61.7 Mb), *TMEM95* (BTA19, 27.6 Mb), and *ARMC3* (BTA13, 24.3 Mb) [27–30]. Although the exact chromosomal positions of these genes differ from precise infertility associated chromosomal regions that have been identified either by QTL studies, GWAS using SCR as parameter, or whole exome sequencing, they are located on the same chromosomes [24,26,41]. Regarding the location of *ABHD16B* on BTA13, it is noteworthy that not only *ARMC3* is located on the same chromosome, but also QTLs for percentage of normal sperms (68.18 cM), male fertility (43.76 cM), and non-return rate (EBV) (85.19 cM) have been mapped to BTA13 [42–44]. One region explaining roughly 0.6% of the genetic variance of SCR was detected on BTA13 from position 58,456,868–59,951,247 harboring two potential candidate genes for male fertility, i.e., *CTCFL* and *SPO11* [45]. This region is located approximately 4 Mb downstream of *ABHD16B*. The closest SNP identified by whole exome sequencing was located on BTA13 at position 53,691,419 within the *SIRPA* gene only 738,396 bp upstream of the nonsense variant in *ABHD16B*. In a further GWAS using a much larger dataset (11.5 k Holstein bulls) and higher density SNP chip (about 300 k), five markers with marked dominance effects were detected, one of them being located on BTA13 (13:g.60263194A>C; rs41701032) [46]. Hence, the molecular genetic data that are published elsewhere are well in agreement with our findings.

The identification of *ABHD16B* as an associated causative gene for bull infertility also allowed for us to elucidate its biochemical function. Except that *ABHD16B* belongs to a large protein superfamily of catalytic enzymes harboring an α/β -hydrolase domain and is predominantly expressed in Leydig cells of the testis, nothing was known regarding its biochemical or physiological function so far [31]. The methylation of *ABHD16B* was reported to be associated with chronic obstructive pulmonary disease (COPD) and aberrant methylation patterns were identified in infertile man [32,36]. In proteome studies, 11 members of the ABHD family (1, 2, 5, 6, 10, 11, 12, 13, 14B, 16A, 17A, 17B) have been detected in testis or spermatozoa [47,48]. Human *ABHD2* participates in sperm hyperactivation as a lipid hydrolase through depleting endocannabinoid 2-arachidonoylglycerol (2-AG), an inhibitor of sperm calcium channel (CatSper) [49]. However, in most cases, their exact role remains elusive.

Our data show that ABHD16B is involved in lipid biosynthesis of DAGs. According to the ABHD16B molecular structure, it is supposed to participate in lipid metabolism, like other ABHD family members, which could contribute to sperm maturation. Sperm lipidomics of two different genotypes was performed to confirm this. Sperm lipid composition changes during their maturation through the epididymis, the percentage of SM in spermatozoa increased [50,51]. SM is synthesized by the combination of Cer and phosphorylcholine from PC. During this reaction, DAG is produced as a by-product [52]. Our data demonstrated that SM and DAG were significantly decreased, while PC and Cer were significantly increased in heterozygous spermatozoa. This implies that ABHD16B might be involved in the lipid biosynthesis of DAG, which influences SM synthesis in the later process. On the other hand, increased levels of PCs in heterozygous sperms could also result from an inhibited degradation from PC to DAG and phosphorylcholine. Cer increased correspondingly without enough phosphorylcholine combined to synthesize SM.

The presence of ABHD16B in the epididymis, as shown by IHC, suggests a role in the lipid metabolism of DAG and SM during sperm maturation. DAG also influences the synthesis of 2-AG, which is an inhibitor of sperm calcium channel (CatSper) preventing sperm hyperactivation. DAG is hydrolyzed to 2-AG by diacylglycerol lipase (DAGL) and, hence, decreased DAG levels in homozygous carrier sperms could result in an insufficient amount of 2-AG leading to a premature capacitation [53]. Furthermore, this effect could be enhanced by the lack of SM. On the other hand, the accumulation of PC and Cer could also interfere with the fertilization capacity. For instance, increased PC concentrations in chicken sperms were reported to be negatively associated with fertility during aging [54]. It has also been observed that imbalanced lipid homeostasis of PC and SM caused sperm membrane instability and infertility in knockout mice [55]. The final sperm lipid composition is formed during epididymal maturation, which results in a decreased amount of cholesterol, PS, CL, PE, and PI, and an increase in PC and DAG. The amount of PI, PC, and DAG was significantly different between the wild type and heterozygous variant spermatozoa (Figure 6A), indicating a potential role of ABHD16B in sperm maturation.

Another impact of ABHD16B on lipid metabolism can be seen in the increased PC:PE ratio. Abnormal PC:PE ratios affect membrane permeability, fluidity, and integrity [56,57]. In cells that have abundant unsaturated fatty acids, such as spermatozoa, LPC is normally regarded as a marker of sperm membrane quality and oxidative stress. The increase of LPC content in the deteriorated membrane of spermatozoa indicates affected acrosome reaction, and an increased ratio between LPC 22:6 and PC 16:0/22:6 was observed in human spermatozoa with impaired membrane [58]. The ratio was also significantly enhanced in the heterozygous samples analysed here. Furthermore, LPC 22:6 is a reliable marker of spermatozoa lipid oxidation [59]. A significant increased concentration of LPC 22:6 was also found in heterozygous carrier samples. This could result in a higher oxidized state or membrane damaged level in contrast to wild type sperms. The lipidomics analysis clearly showed that the loss of ABHD16B function has a profound effect on sperm plasma membrane lipid composition. Therefore, in analogy with experiments in humans and mice, it can be hypothesized that the altered lipid composition of the *ABHD16B* homozygous carrier sperms interferes with the fertilization ability.

4. Materials and Methods

4.1. Ethical Statement

EDTA blood samples of cattle were taken for routine parentage control exclusively by local veterinarians. The Lower Saxony State Office for Consumer Protection and Food Safety approved the collection of samples (33.19-42502-05-17A196), according to §8a Abs. 1 Nr. 2 of the German Animal Protection Law.

4.2. Genome Wide Association Analysis (GWAS)

The conception ability of sires (non-return rate, NR_{dev}) was calculated based on the latest three daughter proven service-sire age groups (2019: A.I. sires born 2014–2016; 0% deviation). The NR_{dev} is expressed in %-deviation on the original non-return-rate scale. Sires in the breeding population with NR_{dev} values of approx. $\pm 2\%$ are scored as average (for more information see https://www.vit.de/fileadmin/DE/Zuchtwirtschaft/Zws_Bes_eng.pdf). Data of NR_{dev} deviations of service-sires were provided by VIT (<https://www.vit.de/en/>).

For GWAS 279 sires of the current breeding population were randomly chosen as presumably fertile controls. As cases 10 sires (including Tarantino, $NR_{dev} = -29$) with NR_{dev} between -29 and -2 were selected (Table 1). The 289 samples were genotyped while using the Illumina BovineSNP50 or MD BeadChip. The chips were processed on a HiScan SQ and iScan System (Illumina GmbH, Munich, Germany) and raw data were converted using GenomeStudio Software (Illumina GmbH, Munich, Germany). Final reports were imported into SVS 8.8.3 for MacOSX (Golden Helix Inc. Bozeman, MT, USA). Prior to GWAS data were filtered while using a call rate <0.95 , number of alleles >2 , minor allele frequency (MAF) <0.05 , and Fisher's HWE <0.001 (based on controls) as marker dropping criteria. LD pruning was performed with a window size of 100 and increments of 5. R^2 -LD statistics with a threshold of 0.5 while using Cochran-Mantel-Haenszel (CHM) as computation method was applied. After filtering, 38,671 markers remained for further analysis. GWAS was done using a multi-locus mixed model (MLMM) while applying an additive genetic model with correction for male X-chromosomal hemizyosity [60,61]. The associations were regarded as statistically significant above a Bonferroni threshold of $-\log_{10}P = 5.9$ ($p = 0.01$). The associations of markers ($-\log_{10}P$ -value, y-axis) were plotted against their chromosomal positions (UMD3.1.1, x-axis).

4.3. Next Generation Sequencing of Tarantino and Its Parents

Tarantino and its parents were sequenced on a HiSeq2500 System (Illumina GmbH, Munich, Germany), resulting in approx. 10^9 total reads per sample. Low quality (average phred quality < 15) and single reads were removed, resulting in approx. 9.4×10^8 per sample. Mapping to the bovine reference genome sequence (UMD3.1.1) was done while using BWA [62]. PCR duplicates were removed using Picard (<http://broadinstitute.github.io/picard/>). After read mapping, alignment and refinement approx. 7.8×10^8 reads remained per sample, corresponding to an average depth of coverage of approx. 46x (mean insert size 360 bp). A total of 9,315,126 SNPs and 1,439,972 indels were called using GATK Haplotype Caller [63]. SNP & Variation Suite 8.8.3 (Golden Helix Inc., Bozeman, MT, USA) was used for further analysis. SNPs and indels were set to missing with read depth ≤ 10 , genotype quality ≤ 15 , alt read ratios for Ref_Ref ≥ 0.15 , Ref_Alt outside 0.3 and 0.7, Alt_Alt ≤ 0.85 , and according to their inheritance pattern (Tarantino = Alt_Alt, parents Alt_Ref). After this filtering, 307,898 SNPs and 604 indels remained. A final filtering was done while using SNPs and indels only in annotated and verified mRNA transcripts, including splice donor and acceptor distances of 2 bp, splice region exonic distances of 3 bp and splice region intronic distances of 8 bp, resulting in 78,472 SNPs and 125 indels.

4.4. Genotyping of SNP rs468948776 (ABHD16B)

The nonsense variant in *ABHD16B* was genotyped while using fluorescence resonance energy transfer (FRET) analysis on a LightCycler 480 (Roche Life Science, Mannheim, Germany). The DNA concentrations were measured using NanoDrop ND-1000 spectrophotometer (PEQLAB Biotechnologie GmbH, Erlangen, Germany). Conventional PCR primers were designed using the online program Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>). The FRET primers were designed with MeltCalc Software [64,65]. Table 2 lists FRET primers and probes.

Table 2. Fluorescence resonance energy transfer (FRET) primers and probes used for genotyping of *ABHD16B* variant.

Gene	Primer Name	Sequence (5'→3')	Probe Name	Sequence (5'→3')
<i>ABHD16B</i>	ABHD16B_FRET_f	ACCCGGGCTTCGGGGGCAGC	ABHD16B_FRET_Pro	[Cy5]CGTTCCTCAGCATGATG[Phos]
	ABHD16B_FRET_r	GCGTACTTGACCACCACGTC	ABHD16B_FRET_Anc	GGGGCAGCACGGGCG[Flc]

SNP rs468948776 (*ABHD16B*) was amplified in a total volume of 25 μ L, including 20 ng DNA, 10 μ mol forward and reverse primer each, 10 μ mol probe and anchor (Sigma-Aldrich, Taufkirchen, Germany) each, 1 \times GC-RICH solution, 1 \times PCR reaction buffer (including 20 mM $MgCl_2$), 100 μ mol dNTPs and FastStart Taq Polymerase (1U; Qiagen, Hilden Germany) for 34 cycles at 95 °C for 15 S, 60 °C for 20 S, and 72 °C for 20 S. The melting curves were done using the following program: 95 °C for 30 S, 37 °C for 30 S, 95 °C continuous acquisition mode (2/°C), ramp rate 0.29 °C/S, followed by 37 °C for 30 S.

4.5. Western Blotting

Immunoblotting on cryopreserved semen specimens of one wild type (G_G), one heterozygous carrier (G_A), two homozygous affected (A_A; Tarantino, C_a), and testis, muscle, and liver samples of wild type bulls were prepared. The semen samples of sire C_a were provided from the safety-inventory of a AI station. Human ABHD16B over-expression lysate (NM_080622, OriGene, Rockville, MD, USA) was used as a positive control. Frozen semen samples were thawed at 37 °C in a water bath for 30 S, followed by 3 \times washes with phosphate-buffered saline (PBS; Invitrogen/ThermoFisher Scientific) and lysed in cold RIPA buffer (Sigma, R0278, St. Louis, MO, USA). Protease inhibitor (Roche, Cat. No.04693159001, Mannheim, Germany) and phosphatase inhibitor (Roche, Cat. No. 04906845001, Germany) were added to RIPA buffer in advance. The samples were incubated for 1 h at 4 °C and centrifuged at 16,000 \times g for 20 min. at 4 °C. An additional homogenization of tissue samples with MagNA lyser green beads (Roche Life Science, Mannheim, Product No. 03358941001, Germany) was carried out followed by an incubation for 2 h at 4 °C and then centrifuged at 16,000 \times g for 20 min. at 4 °C. Protein quantification was performed by Bradford method with the dye reagent concentrate (Bio-Rad, Cat. No. 5000006, Munich, Germany).

After denaturation (10 min. at 70 °C) in LDS sample buffer with 5% 2-mercaptoethanol, equal amounts of protein were loaded to SDS-PAGE (8% Bis-Tris Plus gel, ThermoFisher Scientific, Cat. No. NW00087BOX, USA). After electrophoresis at 15V for 1h, the proteins were transferred onto nitrocellulose membranes (Sigma, Cat. No. 10600098, Germany) with semi-dry blotter (Brenzel Bioanalytik, Lahntal, Germany). Membranes were blocked with 5% non-fat dry milk in TBS-T (0.1% Tween) overnight at 4 °C and then incubated with primary antibodies for 1 h at room temperature, followed by incubation with the secondary antibodies at room temperature for 1 h. Subsequently, the membranes were incubated with an ECL detection reagent (GE Healthcare, Product No. RPN2109, Little Chalfont, UK) and then exposed to X-ray films (GE Healthcare, Product No. 28906836, Tokyo, Japan) for detection.

A customized bovine ABHD16B primary antibody, affinity purification PAC-DFR (Davids Biotechnologie GmbH, Regensburg, Germany, 1 μ g/mL dilution) was used. Goat Anti-Mouse IgG (H + L)-HRP (Bio-Rad, Munich, Germany; 1:10,000 dilution) and Goat Anti-Rabbit IgG (H + L)-HRP (Bio-Rad, Germany; 1:10,000 dilution) were the secondary antibodies. Anti- α -Tubulin (Sigma, T9026; 1: 2500 dilution) was used as the loading control.

The quantification of ABHD16B Western blots was done using ImageJ 1.52k software [66]. Areas under curve of ABHD16B specific bands were determined for liver, muscle, testis and spermatozoa (G_G). Relative expression ratios (%) were calculated with α -tubulin as the internal standard and plotted as Box and Whisker plot.

4.6. Immunohistochemistry of Testes

Testicular and epididymal tissues were obtained from freshly slaughtered wild type Holstein cattle and they were immediately fixed in 4 % formaldehyde for 48 h. Immunohistochemistry (IHC) was performed on paraffin-embedded sections, including testicular parenchyma, as well as ductuli efferentes, epididymal head, corpus, and tail with efferent ducts and epididymal duct, respectively. The primary polyclonal antibody was directed against the PAC-ARK peptide and it was generated in the rabbit according to standard protocols (Davids Biotechnologie GmbH, Regensburg, Germany). IHC was performed in an automated immunostaining system (Discovery XT, Roche Diagnostics GmbH, Mannheim, Germany) at a dilution of 1:1000 while using the SABC (streptavidin-biotin-complex) method, mild EDTA (ethylenediaminetetraacetic acid) pretreatment, and DAB (diaminobenzidine tetrahydrochloride) for signal detection (DAB Map Kit, Roche Diagnostics GmbH, Mannheim, Germany). A rabbit IgG isotype control (ABIN3023746, antibodies-online GmbH, Aachen, Germany) was included at the same concentration as the primary antibody for confirmation of primary antibody specificity. Additionally, pure antibody diluent instead of primary antibody was applied to the control sections for an evaluation of non-specific binding of the secondary antibody.

4.7. Lipidomics of Wild Type and Heterozygous Spermatozoa

4.7.1. Semen Collection for Lipidome Analysis

Wild type and heterozygous fresh semen samples were prepared for lipidome analysis. Three independent fresh ejaculates were collected from a heterozygous bull and five technical replicates were produced by dilution from each sample. Wild type semen samples were flushed from the epididymal tail of four unrelated bulls and a total of 15 technical replicates were generated by dilution. In the epididymal tail, spermatozoa are matured and the lipid composition is equivalent to ejaculated spermatozoa [67,68]. The samples were washed twice in Dulbecco's phosphate-buffered saline (D-PBS) without magnesium and calcium and centrifuged at $1000\times g$ for 5 min. at 4 °C. The cells were resuspended in D-PBS to a final concentration of approximately three million–eight million cells/mL. Cell density was determined in an improved Neubauer counting chamber (Marienfeld GmbH, Lauda-Königshofen, Germany).

4.7.2. Lipid Extraction for Mass Spectrometry Lipidomics

Mass spectrometry-based lipid analysis was performed by Lipotype GmbH (Dresden, Germany), as described [69]. The lipids were extracted while using a two-step chloroform/methanol procedure [70]. The samples were spiked with internal lipid standard mixture containing: cardiolipin 16:1/15:0/15:0/15:0 (CL), ceramide 18:1/2/17:0 (Cer), diacylglycerol 17:0/17:0 (DAG), hexosylceramide 18:1/2/12:0 (HexCer), lyso-phosphatidate 17:0 (LPA), lyso-phosphatidylcholine 12:0 (LPC), lyso-phosphatidylethanolamine 17:1 (LPE), lyso-phosphatidylglycerol 17:1 (LPG), lyso-phosphatidylinositol 17:1 (LPI), lyso-phosphatidylserine 17:1 (LPS), phosphatidate 17:0/17:0 (PA), phosphatidylcholine 17:0/17:0 (PC), phosphatidylethanolamine 17:0/17:0 (PE), phosphatidylglycerol 17:0/17:0 (PG), phosphatidylinositol 16:0/16:0 (PI), phosphatidylserine 17:0/17:0 (PS), cholesterol ester 20:0 (CE), sphingomyelin 18:1/2/12:0/0 (SM), and triacylglycerol 17:0/17:0/17:0 (TAG). After extraction, the organic phase was transferred to an infusion plate and dried in a speed vacuum concentrator. First step dry extract was re-suspended in 7.5 mM ammonium acetate in chloroform/methanol/propanol (1:2:4, V:V:V) and second step dry extract in 33% ethanol solution of methylamine in chloroform/methanol (0.003:5:1; V:V:V). All liquid handling steps were performed while using Hamilton Robotics STARlet robotic platform with the Anti Droplet Control feature for organic solvents pipetting.

4.7.3. MS Data Acquisition

The samples were analyzed by direct infusion on a QExactive mass spectrometer (Thermo Scientific, Osterode am Harz, Germany) equipped with a TriVersa NanoMate ion source (Advion

Biosciences, Ithaca, NY, USA). Samples were analyzed in both positive and negative ion modes with a resolution of $Rm/z=200=280000$ for MS and $Rm/z=200=17500$ for MSMS experiments, in a single acquisition. MSMS was triggered by an inclusion list that encompasses corresponding MS mass ranges scanned in 1 Da increments [71]. MS and MSMS data were both combined to monitor CE, DAG, and TAG ions as ammonium adducts; PC, PC O⁻, as acetate adducts; and, CL, PA, PE, PE O⁻, PG, PI, and PS as deprotonated anions. MS only was used to monitor LPA, LPE, LPE O⁻, LPI, and LPS as deprotonated anions; Cer, HexCer, SM, LPC, and LPC O⁻ as acetate adducts.

4.7.4. Data Analysis and Post-Processing

The data were analyzed with in-house developed lipid identification software based on LipidXplorer [72,73]. Data post-processing and normalization were performed while using an in-house developed data management system. Only lipid identifications with a signal-to-noise ratio >5, and a signal intensity five-fold higher than in corresponding blank samples were considered for further data analysis.

The total lipid amount occurring in each sperm sample were pre-tested to ensure that optimal amounts are used to achieve the greatest analysis quality and result comparability, despite the broad dynamic range of our analytical methods. Afterwards, the initially detected total lipid amount per sample was normalized to one-million sperms. A significant difference between normalized total lipid amount of wild type and heterozygous samples was analyzed.

A 70% occupational threshold was applied for data filter, valid data in more than 10 samples for each genotype were selected, NAs were replaced with zeros. Afterwards, lipid data that were present in both genotypes were chosen for further analysis. In total, 16 lipid classes with 144 lipid species were evaluated, and the data were analyzed in terms of lipid class and species separately. Shapiro–Wilk Test was used for normal distribution detection [74]. Significant difference analysis was performed with Mann–Whitney U-test or two-tailed t-test, depending on the normal distribution results by SPSS 16.0. PC:PE and (LPC 22:6):(PC 16:0_22:6) ratios were also checked for significant difference. Benjamini & Hochberg method was used for *p*-value adjustment of multiple testing [75] with R version 3.5.1, $p_{(BH)} < 0.05$ were considered to be statistically significant (Table S1). The comparison results of lipid classes and 10 most significantly changed lipid species between wild-type and rs468948776 heterozygous samples were demonstrated in histograms (data are presented with mean \pm standard error of mean). 144 lipid species were plotted with Y-axis of adjusted values ($-\log_{10}p_{BH}$) against X-axis of \log_2 fold change (heterozygous vs wild type).

5. Conclusions

We have identified a nonsense mutation in the bovine *ABHD16B* gene as a potential causative protein-altering variant for male infertility in Holstein cattle. This made it possible to elucidate the so far unknown physiological and biochemical role of ABHD16B in lipid biosynthesis, spermatogenesis, and fertilization. Our findings could also have implications on further elucidating a novel genetic cause for human male infertility, due to the fact that a number of deleterious variants, e.g., missense, frameshift, indels and one stop-gain variant in the human *ABHD16B* gene have been reported to the human ENSEMBL database.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/21/2/627/s1>.

Author Contributions: S.S., F.X., M.B., S.B., T.M., M.H. and C.W. conducted the experiments. W.W. collected samples. S.S., F.X., E.S. and B.B. analyzed the data. B.B. planned and supervised the project. B.B. and S.Z. supervised S.S. and F.X. S.S., F.X., M.B. and B.B. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors are grateful to S. Pach for expert technical assistance. Masterrind is thanked for providing semen samples of sires. The Association for Bioeconomy Research (FBF), German Livestock Association (BRS) and IT-Solutions for Animal Production (VIT) are thanked for support and pedigree data. S. Shan and

F. Xu are fellows of the China Scholarship Council (CSC). We acknowledge support by the German Research Foundation and the Open Access Publications of the University of Goettingen.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sakaguchi, M. Practical aspects of the fertility of dairy cattle. *J. Reprod. Dev.* **2011**, *57*, 17–33. [\[CrossRef\]](#)
2. Veerkamp, R.F.; Beerda, B. Genetics and genomics to improve fertility in high producing dairy cows. *Theriogenology* **2007**, *68* (Suppl. 1), S266–S273. [\[CrossRef\]](#)
3. Lucy, M.C. Fertility in high-producing dairy cows: Reasons for decline and corrective strategies for sustainable improvement. *Soc. Reprod. Fertil. Suppl.* **2007**, *64*, 237–254. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Olynk, N.J.; Wolf, C.A. Economic analysis of reproductive management strategies on US commercial dairy farms. *J. Dairy Sci.* **2008**, *91*, 4082–4091. [\[CrossRef\]](#)
5. Harstine, B.R.; Utt, M.D.; DeJarnette, J.M. Review: Integrating a semen quality control program and sire fertility at a large artificial insemination organization. *Animal* **2018**, *1*–12. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Yaniz, J.L.; Soler, C.; Alquezar-Baeta, C.; Santolaria, P. Toward an integrative and predictive sperm quality analysis in *Bos taurus*. *Anim. Reprod. Sci.* **2017**, *181*, 108–114. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Morrell, J.M.; Nongbua, T.; Valeanu, S.; Lima Verde, I.; Lundstedt-Enkel, K.; Edman, A.; Johannisson, A. Sperm quality variables as indicators of bull fertility may be breed dependent. *Anim. Reprod. Sci.* **2017**, *185*, 42–52. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Puglisi, R.; Gaspa, G.; Balduzzi, D.; Severgnini, A.; Vanni, R.; Macciotta, N.; Galli, A. Genomewide analysis of bull sperm quality and fertility traits. *Reprod. Domest. Anim.* **2016**, *51*, 840–843. [\[CrossRef\]](#)
9. Kastelic, J.P. Understanding and evaluating bovine testes. *Theriogenology* **2014**, *81*, 18–23. [\[CrossRef\]](#)
10. Amann, R.P.; DeJarnette, J.M. Impact of genomic selection of AI dairy sires on their likely utilization and methods to estimate fertility: A paradigm shift. *Theriogenology* **2012**, *77*, 795–817. [\[CrossRef\]](#)
11. Kealey, C.G.; MacNeil, M.D.; Tess, M.W.; Geary, T.W.; Bellows, R.A. Genetic parameter estimates for scrotal circumference and semen characteristics of Line 1 Hereford bulls. *J. Anim. Sci.* **2006**, *84*, 283–290. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Feugang, J.M.; Rodriguez-Osorio, N.; Kaya, A.; Wang, H.; Page, G.; Ostermeier, G.C.; Topper, E.K.; Memili, E. Transcriptome analysis of bull spermatozoa: Implications for male fertility. *Reprod. Biomed. Online* **2010**, *21*, 312–324. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Selvaraju, S.; Parthipan, S.; Somashekar, L.; Kolte, A.P.; Krishnan Binsila, B.; Arangasamy, A.; Ravindra, J.P. Occurrence and functional significance of the transcriptome in bovine (*Bos taurus*) spermatozoa. *Sci. Rep.* **2017**, *7*, 42392. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Lalancette, C.; Thibault, C.; Bachand, I.; Caron, N.; Bissonnette, N. Transcriptome analysis of bull semen with extreme nonreturn rate: Use of suppression-subtractive hybridization to identify functional markers for fertility. *Biol. Reprod.* **2008**, *78*, 618–635. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Samanta, L.; Swain, N.; Ayaz, A.; Venugopal, V.; Agarwal, A. Post-Translational Modifications in sperm Proteome: The Chemistry of Proteome diversifications in the Pathophysiology of male factor infertility. *Biochim. Biophys. Acta* **2016**, *1860*, 1450–1465. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Mohanty, G.; Swain, N.; Samanta, L. Sperm Proteome: What Is on the Horizon? *Reprod. Sci.* **2015**, *22*, 638–653. [\[CrossRef\]](#)
17. Legare, C.; Droit, A.; Fournier, F.; Bourassa, S.; Force, A.; Cloutier, F.; Tremblay, R.; Sullivan, R. Investigation of male infertility using quantitative comparative proteomics. *J. Proteome Res.* **2014**, *13*, 5403–5414. [\[CrossRef\]](#)
18. Gan, H.; Cai, T.; Lin, X.; Wu, Y.; Wang, X.; Yang, F.; Han, C. Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Mol. Cell. Proteom.* **2013**, *12*, 1144–1157. [\[CrossRef\]](#)
19. Panner Selvam, M.K.; Agarwal, A. Update on the proteomics of male infertility: A systematic review. *Arab J. Urol.* **2018**, *16*, 103–112. [\[CrossRef\]](#)
20. Abdollahi-Arpanahi, R.; Morota, G.; Penagaricano, F. Predicting bull fertility using genomic data and biological information. *J. Dairy Sci.* **2017**, *100*, 9656–9666. [\[CrossRef\]](#)
21. Suchocki, T.; Szyda, J. Genome-wide association study for semen production traits in Holstein-Friesian bulls. *J. Dairy Sci.* **2015**, *98*, 5774–5780. [\[CrossRef\]](#) [\[PubMed\]](#)

22. Taylor, J.F.; Schnabel, R.D.; Sutovsky, P. Review: Genomics of bull fertility. *Animal* **2018**, *12*, s172–s183. [CrossRef] [PubMed]
23. Penagaricano, F.; Weigel, K.A.; Khatib, H. Genome-wide association study identifies candidate markers for bull fertility in Holstein dairy cattle. *Anim. Genet.* **2012**, *43* (Suppl. 1), 65–71. [CrossRef] [PubMed]
24. Han, Y.; Penagaricano, F. Unravelling the genomic architecture of bull fertility in Holstein cattle. *BMC Genet.* **2016**, *17*, 143. [CrossRef] [PubMed]
25. Ogorevc, J.; Dovc, P.; Kunej, T. Comparative genomics approach to identify candidate genetic loci for male fertility. *Reprod. Domest. Anim.* **2011**, *46*, 229–239. [CrossRef] [PubMed]
26. Whiston, R.; Finlay, E.K.; McCabe, M.S.; Cormican, P.; Flynn, P.; Cromie, A.; Hansen, P.J.; Lyons, A.; Fair, S.; Lonergan, P.; et al. A dual targeted beta-defensin and exome sequencing approach to identify, validate and functionally characterise genes associated with bull fertility. *Sci. Rep.* **2017**, *7*, 12287. [CrossRef]
27. Pausch, H.; Kolle, S.; Wurmser, C.; Schwarzenbacher, H.; Emmerling, R.; Jansen, S.; Trottmann, M.; Fuerst, C.; Gotz, K.U.; Fries, R. A nonsense mutation in TMEM95 encoding a nondescript transmembrane protein causes idiopathic male subfertility in cattle. *PLoS Genet.* **2014**, *10*, e1004044. [CrossRef]
28. Pausch, H.; Venhoranta, H.; Wurmser, C.; Hakala, K.; Iso-Touru, T.; Sironen, A.; Vingborg, R.K.; Lohi, H.; Soderquist, L.; Fries, R.; et al. A frameshift mutation in ARMC3 is associated with a tail stump sperm defect in Swedish Red (*Bos taurus*) cattle. *BMC Genet.* **2016**, *17*, 49. [CrossRef]
29. Fernandez-Fuertes, B.; Laguna-Barraza, R.; Fernandez-Gonzalez, R.; Gutierrez-Adan, A.; Blanco-Fernandez, A.; O'Doherty, A.M.; Di Fenza, M.; Kelly, A.K.; Kolle, S.; Lonergan, P. Subfertility in bulls carrying a nonsense mutation in transmembrane protein 95 is due to failure to interact with the oocyte vestments. *Biol. Reprod.* **2017**, *97*, 50–60. [CrossRef]
30. Dai, L.; Xu, Y.; Yu, W.; Liu, S.; Gao, Y.; Zhang, L.; Yuan, B.; Chen, J.; Ma, T.; Zhang, J. Naturally occurring genetic mutations in the 5'-upstream regulatory region of bovine FSHB generate a novel cis-regulatory element that affects its expression. *Anim. Genet.* **2015**, *46*, 693–696. [CrossRef]
31. Lord, C.C.; Thomas, G.; Brown, J.M. Mammalian alpha beta hydrolase domain (ABHD) proteins: Lipid metabolizing enzymes at the interface of cell signaling and energy metabolism. *Biochim. Biophys. Acta* **2013**, *1831*, 792–802. [CrossRef] [PubMed]
32. Wan, E.S.; Qiu, W.; Baccarelli, A.; Carey, V.J.; Bacherman, H.; Rennard, S.I.; Agusti, A.; Anderson, W.H.; Lomas, D.A.; DeMeo, D.L. Systemic steroid exposure is associated with differential methylation in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **2012**, *186*, 1248–1255. [CrossRef] [PubMed]
33. Kurima, K.; Ebrahim, S.; Pan, B.; Sedlacek, M.; Sengupta, P.; Millis, B.A.; Cui, R.; Nakanishi, H.; Fujikawa, T.; Kawashima, Y.; et al. TMC1 and TMC2 Localize at the Site of Mechanotransduction in Mammalian Inner Ear Hair Cell Stereocilia. *Cell Rep.* **2015**, *12*, 1606–1617. [CrossRef] [PubMed]
34. Pan, B.; Geleoc, G.S.; Asai, Y.; Horwitz, G.C.; Kurima, K.; Ishikawa, K.; Kawashima, Y.; Griffith, A.J.; Holt, J.R. TMC1 and TMC2 are components of the mechanotransduction channel in hair cells of the mammalian inner ear. *Neuron* **2013**, *79*, 504–515. [CrossRef] [PubMed]
35. Uhlen, M.; Fagerberg, L.; Hallstrom, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. *Science* **2015**, *347*, 1260419. [CrossRef] [PubMed]
36. Carrell, D.T.; Cairns, B.; Aston, K.I.; Jenkins, T.; Smith, A.D.; Uren, P.J.; Horsager, A. Methods of Identifying Male Fertility Status and Embryo Quality. 2019. Available online: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2017024311&_cid=P22-K5CGNG-22960-1. (accessed on 16 January 2020).
37. Moghbelinejad, S.; Mozdarania, H.; Ghoraeian, P.; Asadi, R. Basic and clinical genetic studies on male infertility in Iran during 2000–2016: A review. *Int. J. Reprod. Biomed. (Yazd)* **2018**, *16*, 131–148. [CrossRef]
38. Krausz, C.; Riera-Escamilla, A. Genetics of male infertility. *Nat. Rev. Urol.* **2018**, *15*, 369–384. [CrossRef]
39. Bracke, A.; Peeters, K.; Punjabi, U.; Hoogewijs, D.; Dewilde, S. A search for molecular mechanisms underlying male idiopathic infertility. *Reprod. Biomed. Online* **2018**, *36*, 327–339. [CrossRef]
40. Tüttelmann, F.; Ruckert, C.; Ropke, A. Disorders of spermatogenesis: Perspectives for novel genetic diagnostics after 20 years of unchanged routine. *Med. Genet.* **2018**, *30*, 12–20. [CrossRef]
41. Li, G.; Penagaricano, F.; Weigel, K.A.; Zhang, Y.; Rosa, G.; Khatib, H. Comparative genomics between fly, mouse, and cattle identifies genes associated with sire conception rate. *J. Dairy Sci.* **2012**, *95*, 6122–6129. [CrossRef]

42. Blaschek, M.; Kaya, A.; Zwald, N.; Memili, E.; Kirkpatrick, B.W. A whole-genome association analysis of noncompensatory fertility in Holstein bulls. *J. Dairy Sci.* **2011**, *94*, 4695–4699. [[CrossRef](#)] [[PubMed](#)]
43. Fortes, M.R.; Reverter, A.; Kelly, M.; McCulloch, R.; Lehnert, S.A. Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. *Andrology* **2013**, *1*, 644–650. [[CrossRef](#)] [[PubMed](#)]
44. Hoglund, J.K.; Sahana, G.; Guldbrandtsen, B.; Lund, M.S. Validation of associations for female fertility traits in Nordic Holstein, Nordic Red and Jersey dairy cattle. *BMC Genet.* **2014**, *15*, 8. [[CrossRef](#)] [[PubMed](#)]
45. Nicolini, P.; Amorin, R.; Han, Y.; Penagaricano, F. Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. *BMC Genet.* **2018**, *19*, 14. [[CrossRef](#)]
46. Nani, J.P.; Rezende, F.M.; Penagaricano, F. Predicting male fertility in dairy cattle using markers with large effect and functional annotation data. *BMC Genom.* **2019**, *20*, 258. [[CrossRef](#)]
47. Wei, W.; Luo, W.; Wu, F.; Peng, X.; Zhang, Y.; Zhang, M.; Zhao, Y.; Su, N.; Qi, Y.; Chen, L.; et al. Deep Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer. *J. Proteome Res.* **2016**, *15*, 3988–3997. [[CrossRef](#)]
48. Chalmel, F.; Lardenois, A.; Evrard, B.; Mathieu, R.; Feig, C.; Demougin, P.; Gattiker, A.; Schulze, W.; Jegou, B.; Kirchhoff, C.; et al. Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum. Reprod.* **2012**, *27*, 3233–3248. [[CrossRef](#)]
49. Miller, M.R.; Mannowetz, N.; Iavarone, A.T.; Safavi, R.; Gracheva, E.O.; Smith, J.F.; Hill, R.Z.; Bautista, D.M.; Kirichok, Y.; Lishko, P.V. Unconventional endocannabinoid signaling governs sperm activation via the sex hormone progesterone. *Science* **2016**, *352*, 555–559. [[CrossRef](#)]
50. Rejraji, H.; Sion, B.; Prensier, G.; Carreras, M.; Motta, C.; Frenoux, J.-M.; Vericel, E.; Grizard, G.; Vernet, P.; Drevet, J.R. Lipid remodeling of murine epididymosomes and spermatozoa during epididymal maturation. *Biol. Reprod.* **2006**, *74*, 1104–1113. [[CrossRef](#)]
51. Martínez, P.; Morros, A. Membrane lipid dynamics during human sperm capacitation. *Front. Biosci.* **1996**, *1*, d103–d117.
52. Merrill, A.; Sweeley, C.C. Sphingolipids: Metabolism and cell signalling. *New Compr. Biochem.* **1996**, *31*, 309–339.
53. Fonseca, B.; Costa, M.; Almada, M.; Correia-da-Silva, G.; Teixeira, N. Endogenous cannabinoids revisited: A biochemistry perspective. *Prostaglandins Other Lipid Mediat.* **2013**, *102*, 13–30. [[CrossRef](#)] [[PubMed](#)]
54. Cerolini, S.; Kelso, K.; Noble, R.; Speake, B.; Pizzi, F.; Cavalchini, L. Relationship between spermatozoan lipid composition and fertility during aging of chickens. *Biol. Reprod.* **1997**, *57*, 976–980. [[CrossRef](#)] [[PubMed](#)]
55. Bjorkgren, I.; Gylling, H.; Turunen, H.; Huhtaniemi, I.; Strauss, L.; Poutanen, M.; Sipila, P. Imbalanced lipid homeostasis in the conditional Dicer1 knockout mouse epididymis causes instability of the sperm membrane. *FASEB J.* **2015**, *29*, 433–442. [[CrossRef](#)] [[PubMed](#)]
56. Min-Seok, R.; Kawamata, Y.; Nakamura, H.; Ohta, A.; Takagi, M. Isolation and characterization of ECT1 gene encoding CTP: Phosphoethanolamine cytidylyltransferase of *Saccharomyces cerevisiae*. *J. Biochem.* **1996**, *120*, 1040–1047. [[CrossRef](#)] [[PubMed](#)]
57. Li, Z.; Agellon, L.B.; Allen, T.M.; Umeda, M.; Jewell, L.; Mason, A.; Vance, D.E. The ratio of phosphatidylcholine to phosphatidylethanolamine influences membrane integrity and steatohepatitis. *Cell Metab.* **2006**, *3*, 321–331. [[CrossRef](#)]
58. Glander, H.J.; Schiller, J.; Suss, R.; Paasch, U.; Grunewald, S.; Arnhold, J. Deterioration of spermatozoal plasma membrane is associated with an increase of sperm lyso-phosphatidylcholines. *Andrologia* **2002**, *34*, 360–366. [[CrossRef](#)]
59. Fuchs, B.; Muller, K.; Goritz, F.; Blottner, S.; Schiller, J. Characteristic oxidation products of choline plasmalogens are detectable in cattle and roe deer spermatozoa by MALDI-TOF mass spectrometry. *Lipids* **2007**, *42*, 991–998. [[CrossRef](#)]
60. Kang, H.M.; Sul, J.H.; Service, S.K.; Zaitlen, N.A.; Kong, S.Y.; Freimer, N.B.; Sabatti, C.; Eskin, E. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **2010**, *42*, 348–354. [[CrossRef](#)]
61. Vilhjalmsón, B.J.; Nordborg, M. The nature of confounding in genome-wide association studies. *Nat. Rev. Genet.* **2013**, *14*, 1–2. [[CrossRef](#)]

62. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
63. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*. [[CrossRef](#)]
64. Schutz, E.; Von Ahsen, N. Spreadsheet software for thermodynamic melting point prediction of oligonucleotide hybridization with and without mismatches. *Biotechniques* **1999**, *27*, 1218–1222, 1224. [[CrossRef](#)] [[PubMed](#)]
65. Von Ahsen, N.; Oellerich, M.; Armstrong, V.W.; Schutz, E. Application of a thermodynamic nearest-neighbor model to estimate nucleic acid stability and optimize probe design: Prediction of melting points of multiple mutations of apolipoprotein B-3500 and factor V with a hybridization probe genotyping assay on the LightCycler. *Clin. Chem.* **1999**, *45*, 2094–2101. [[PubMed](#)]
66. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [[CrossRef](#)] [[PubMed](#)]
67. Johnson, L.A.; Pursel, V.G.; Gerrits, R.J. Total phospholipid and phospholipid fatty acids of ejaculated and epididymal semen and seminal vesicle fluids of boars. *J. Anim. Sci.* **1972**, *35*, 398–403. [[CrossRef](#)]
68. Angrimani, D.S.R.; Nichi, M.; Losano, J.D.A.; Lucio, C.F.; Veiga, G.A.L.; Franco, M.V.M.J.; Vannucchi, C.I. Fatty acid content in epididymal fluid and spermatozoa during sperm maturation in dogs. *J. Anim. Sci. Biotechnol.* **2017**, *8*, 18. [[CrossRef](#)]
69. Sampaio, J.L.; Gerl, M.J.; Klose, C.; Ejsing, C.S.; Beug, H.; Simons, K.; Shevchenko, A. Membrane lipidome of an epithelial cell line. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1903–1907. [[CrossRef](#)]
70. Ejsing, C.S.; Sampaio, J.L.; Surendranath, V.; Duchoslav, E.; Ekroos, K.; Klemm, R.W.; Simons, K.; Shevchenko, A. Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2136–2141. [[CrossRef](#)]
71. Surma, M.A.; Herzog, R.; Vasilj, A.; Klose, C.; Christinat, N.; Morin-Rivron, D.; Simons, K.; Masoodi, M.; Sampaio, J.L. An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids. *Eur. J. Lipid Sci. Technol.* **2015**, *117*, 1540–1549. [[CrossRef](#)]
72. Herzog, R.; Schuhmann, K.; Schwudke, D.; Sampaio, J.L.; Bornstein, S.R.; Schroeder, M.; Shevchenko, A. LipidXplorer: A software for consensual cross-platform lipidomics. *PLoS ONE* **2012**, *7*, e29851. [[CrossRef](#)] [[PubMed](#)]
73. Herzog, R.; Schwudke, D.; Schuhmann, K.; Sampaio, J.L.; Bornstein, S.R.; Schroeder, M.; Shevchenko, A. A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language. *Genome Biol.* **2011**, *12*, R8. [[CrossRef](#)] [[PubMed](#)]
74. Zorn, C. Shapiro-Wilk test. In *The SAGE Encyclopedia of Social Science Research Methods*; Sage Publications, Inc.: Thousand Oaks, CA, USA, 2004; pp. 1030–1031.
75. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **1995**, *57*, 289–300. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

CHAPTER 3

Christmas disease in a Hovawart family resembling human hemophilia B Leyden is caused by a single nucleotide deletion in a highly conserved transcription factor binding site of the *F9* gene promoter

Bertram Brenig^{† *}, Lilith Steingraber[†], Shuwen Shan, Fangzheng Xu, Marc Hirschfeld, Reiner Andag, Mirjam Spengeler, Elisabeth Dietschi, Reinhard Mischke, and Tosso Leeb

[†] These authors contributed equally to this work.

^{*} Corresponding author

The article was published online in *haematologica*, 104(11), p.2307. in November 2019. The full article can be found online at:

<https://doi.org/10.3324/haematol.2018.215426>

Author contributions:

Shuwen Shan conducted the experiment of electrophoretic mobility shift assay and wrote the manuscript.

Christmas disease in a Hovawart family resembling human hemophilia B Leyden is caused by a single nucleotide deletion in a highly conserved transcription factor binding site of the *F9* gene promoter



Bertram Brenig,^{1*} Lilith Steingraber^{1*}, Shuwen Shan,¹ Fangzheng Xu,¹ Marc Hirschfeld,^{1,2} Reiner Andag,³ Mirjam Spengeler,⁴ Elisabeth Dietschi,⁴ Reinhard Mischke⁵ and Tosso Leeb⁴

¹University of Göttingen, Institute of Veterinary Medicine, Göttingen, Germany;

²Department of Obstetrics and Gynecology, Freiburg University Medical Center, Freiburg, Germany; ³University Medical Center Göttingen, Institute for Clinical Chemistry, Göttingen, Germany; ⁴Institute of Genetics, University of Bern, Bern, Switzerland and

⁵Small Animal Clinic, University of Veterinary Medicine Hannover Foundation, Hannover, Germany

*BB and LS contributed equally to this work.

Haematologica 2019
Volume 104(11):2307-2313

ABSTRACT

Hemophilia B is a classical monogenic, X-chromosomal, recessively transmitted bleeding disorder caused by genetic variants within the coagulation factor IX gene (*F9*). Although hemophilia B has been described in dogs, it has not yet been reported in the Hovawart breed. Here we describe the identification of a Hovawart family transmitting typical signs of an X-linked bleeding disorder. Five males were reported to suffer from recurrent hemorrhagic episodes. A blood sample from one of these males with only 2% of the normal concentration of plasma factor IX together with samples from seven relatives were provided. Next-generation sequencing of the mother and grandmother revealed a single nucleotide deletion in the *F9* promoter. Genotyping of the deletion in 1,298 dog specimens including 720 Hovawarts revealed that the mutant allele was only present in the aforementioned Hovawart family. The deletion is located 73 bp upstream of the *F9* start codon in the conserved overlapping DNA binding sites of hepatocyte nuclear factor 4α (HNF-4α) and androgen receptor (AR). The deletion only abolished binding of HNF-4α, while AR binding was unaffected as demonstrated by electrophoretic mobility shift assay using human HNF-4α and AR with double-stranded DNA probes encompassing the mutant promoter region. Luciferase reporter assays using wild-type and mutated promoter fragment constructs transfected into Hep G2 cells showed a significant reduction in expression from the mutant promoter. The data provide evidence that the deletion in the Hovawart family caused a rare type of hemophilia B resembling human hemophilia B Leyden.

Introduction

Hemophilia B (Christmas disease) is a recessive, X-linked bleeding disorder caused by genetic variants within the clotting factor IX gene (*F9*) resulting in the absence or insufficient levels of factor IX (FIX) in the blood.¹ In humans hemophilia B is also known as the “royal disease” as it was transmitted into several European royal dynasties by Queen Victoria.^{2,3} As of present, 1,113 unique *F9* variants have been described in humans.⁴ The majority of the pathogenic variants are located within exons (n=923) and intronic regions (n=137) of *F9*. Only 33 variants (2.96%) have been described in the 5'-UTR (n=28) and 3'-UTR (n=5) accounting for 2.52% and 0.45% of human pathogenic hemophilia B variants, respectively.⁴

Although the first reports about canine hemophilia B date back to the early 1960s

Correspondence:

BERTRAM BRENIG
bbrenig@gwdg.de

Received: December 26, 2018

Accepted: March 6, 2019.

Pre-published: March 7, 2019.

doi:10.3324/haematol.2018.215426

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/104/11/2307

©2019 Ferrata Storti Foundation

Material published in *Haematologica* is covered by copyright. All rights are reserved to the Ferrata Storti Foundation. Use of published material is allowed under the following terms and conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>. Copies of published material are allowed for personal or internal use. Sharing published material for non-commercial purposes is subject to the following conditions: <https://creativecommons.org/licenses/by-nc/4.0/legalcode>, sect. 3. Reproducing and sharing published material for commercial purposes is not allowed without permission in writing from the publisher.



and this was the first disorder in dogs characterized on the DNA level, data on hemophilia B cases in dogs remain rather scarce compared to data from humans.⁵⁻⁸ For instance, in the Cairn Terrier colony of the Francis Owen Blood Research Laboratory (University of North Carolina, Chapel Hill, NC, USA) a G>A transition (NC_006621.3:g.109,532,018G>A) in exon 8 causing an amino acid exchange (NP_001003323.1:p.Gly418Glu) that resulted in a complete lack of circulating FIX was detected in affected dogs.⁹ Due to a complete deletion of *F9* in a Labrador Retriever, FIX inhibitors were produced after transfusion of canine blood products.¹⁰ In a study of Pit Bull Terrier mixed breed dogs and Airedale Terrier dogs a large deletion of the entire 5' region of *F9* extending to exon 6 was found in the former and a 5 kb insertion disrupting exon 8 was described in the latter.¹¹ As in the Labrador Retriever with hemophilia B, FIX inhibitors were produced in both breeds. A mild form of hemophilia B in German Wirehaired Pointers was caused by a 1.5 kb Line-1 insertion in intron 5 of *F9* at position NC_006621.3:g.109,521,130.¹² Until today, hemophilia B has been described in four mixed-breed dogs and nine dog breeds, i.e. German Shepherd, Lhasa Apso, Labrador Retriever, Rhodesian Ridgeback, Airedale Terrier, Cairn Terrier, Maltese, Mongrel and German Wirehaired Pointer.⁹⁻¹⁷

In the canine cases analyzed so far on the DNA level, mutations have been observed only in exons and introns of *F9*, whereas alterations of the *F9* promoter have not yet been reported. In humans promoter variants have been detected and result in the so-called hemophilia B Leyden characterized by low levels of FIX until puberty, whereas after puberty FIX concentrations rise to almost normal levels.¹⁸⁻²⁰ Since its first description, the genetic background of human hemophilia B Leyden was elucidated by various studies identifying variants in different transcription factor binding sites in the *F9* promoter including the androgen-responsive element (ARE), hepatocyte nuclear factor 4α (HNF4α), one cut homeobox (ONECUT1/2) and CCAAT/enhancing-binding protein α (C/EBPα) binding sites.^{21,22} HNF4α is a liver-enriched member of the nuclear receptor superfamily of ligand-dependent transcription factors and has been associated with several disorders, including diabetes, atherosclerosis, hepatitis, cancer, and hemophilia.²³ Promoter analyses have identified at least 140 genes with HNF4α binding sites. A recent, more detailed analysis using protein binding microarrays identified an additional 1,400 potential binding sites.^{24,25} Hence, HNF4α plays an important role in the regulation of numerous genes especially in the maintenance of many liver-specific functions. Liver-specific HNF4α-null mice have been used to study the involvement of hepatic HNF4α in blood coagulation. In the murine model it was shown that expression of factors V, XI, XII, and XIIIB depends directly on hepatic HNF4α and FIX expression was decreased with significantly prolonged activated partial thromboplastin time (aPTT).²⁶ Ten of the so far identified 28 5'-UTR variants (35.7%) are located within the overlapping binding sites of the androgen receptor (AR) and HNF4α in the human *F9* promoter.^{4,21} Four variants at positions -21, -20 and -19 only affect HNF4α binding and all of them have been shown to cause hemophilia B Leyden.^{19,27-30} The remaining six variants at positions -26, -24 and -23, located in the overlapping region, cause the so-called hemophilia B

Brandenburg.^{31,32} Unlike the classical hemophilia B Leyden, FIX levels in patients with these variants cannot be restored by testosterone-driven AR activity and remain low after puberty with no clinical recovery.^{21,32}

Methods

Animals and genomic DNA isolation

Canine blood and/or hair samples were collected by local veterinarians. The collection of samples was approved by the Lower Saxony State Office for Consumer Protection and Food Safety (33.19-42502-05-15A506) according to §8a Abs. 1 Nr. 2 of the German Animal Protection Law (TierSchG). Blood collected into EDTA and/or hair samples were provided by different Hovawart and dog breeders with written consent from the dogs' owners. DNA was extracted from 30-50 hair roots using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.³³ A salting out procedure was used to obtain DNA from the EDTA blood samples.³⁴ Additional DNA samples deposited with the Institute of Veterinary Medicine were used as controls. All samples were pseudonymized using internal identities.

Next-generation sequencing and genotyping

DNA from animals #4 and #6 was used for next-generation sequencing on an Illumina HiSeq2500. The quality of the fastq-files was analyzed using FastQC 0.11.7.³⁵ Total reads of 1,029,601,630 (#4; sequencing depth 51x) and 1,000,503,256 (#6; sequencing depth 50x) were obtained and mapped to the reference canine *F9* gene (NC_006621.3, region 109,501,341 to 109,533,798; CanFam3.1) using DNASTAR Lasergene Genomics Suite SeqMan NGen 15.2.0 (130).³⁶⁻⁴⁰

Targeted genotyping of the promoter deletion was done by polymerase chain reaction (PCR) amplification with primers cfa_F9_Ex1_F (5'-CCACTGAGGGAGATGGACAC-3') and cfa_F9_Ex1_R (5'-CCCACATGCTGACGACTAGA-3') resulting in a fragment of 328 bp (wildtype) or 327 bp (deletion) spanning the variant position. The resulting PCR products were either directly sequenced on an ABI 3730 Genetic Analyzer (Thermo Fisher Scientific, Basel, Switzerland) or genotypes were determined by restriction fragment length polymorphism analysis after cleavage with *RsaI*. The wildtype allele generated two fragments of 52 bp and 276 bp while the allele with the deletion remained uncut.

Electrophoretic mobility shift assay

For the electrophoretic mobility shift assay, biotin-labeled, double-stranded wildtype (cfa_F9n_wt_Biotin: 5'-CAGAAGTAAAT-ACAGCTCAACTTGTACTTTGGAACAACCTGGTCAACC-3') and mutated (cfa_F9n_mut_Biotin: 5'-CCAGAAGTAAAT-ACAGCTCAACTTGTATTTGGAACAACCTGGTCAACC-3') oligonucleotides were synthesized (Integrated DNA Technologies IDT, Leuven, Belgium) harboring the overlapping HNF4α and AR binding sites (underlined). The position of the deleted C-nucleotide is indicated in bold and italics. Recombinant human HNF4α and human AR overexpression lysate were purchased from Origene Technologies Inc. (Rockville, MD, USA).

DNA was detected using the Chemiluminescent Nucleic Acid Detection Module Kit (Thermo Scientific, USA) with minor modifications, i.e. membranes were incubated for 1 min in the substrate working solution.

Luciferase assay

pGL3 Luciferase Reporter Vectors (pGL3-Basic, pGL3-Control)

were used for the luciferase assay (Promega, Mannheim, Germany). The wildtype *F9* promoter fragment (971 bp wildtype) was generated by PCR using primers cfa9_HindIII_F_neu (5'-CGTAGACTTAGCACTGTTCAAAGCTTCACACACACAGTTCTTAAAT-3') and cfa9_HindIII_R_neu (5'-ATGGCTAGCAACCGTCTAAGAAGCTTAATTGTGCAAGGAGCAAGG-3'). The mutated *F9* promoter fragment (970 bp) was generated by PCR using primers cfa9_HindIII_F (5'-ATCGTCAAAGCTTCACACACACAGTTCTTAAAT-3') and cfa9_HindIII_R (5'-CGTACGAAGCTTAATTGTGCAAGGAGCAAGG-3'). For cloning into the *Hind*III restriction site of pGL3, primers were designed with an unspecific random 5'-tag (italics) followed by a *Hind*III restriction site (underlined) (*Online Supplementary Figure S1*). DNA from female carrier #6 served as a template for amplification. Low expression levels of C/EBP in Hep G2 cells were complemented by co-transfection of a C/EBP α expression vector.²² Data are presented as relative response ratios.⁴¹ A Mann-Whitney U test was used to determine statistical significance. Values were considered statistically significant when $P < 0.05$ (weakly significant), $P < 0.01$ (medium-strength significance) and $P < 0.001$ (strongly significant).

Other methods

Further details of the study methods are given in the *Online Supplementary Appendix*.

Results

Hemophilias are rare diseases in dogs and hence it was rather coincidental that a case in a Hovawart (#3, Figure 1) was reported to us. With the reconstruction of the pedigree using the online dog breed database and pedigree data of individual dogs provided by the owners it was possible to trace the disease back to the female carrier #39 (Figure 1, *Online Supplementary Figure S2*).⁴² In the studied family the hemophilia was transmitted to animals #19, #4 and #6. Bitch #19 had one litter with three hemophilic males (#48, #51, #53). Bitches #4 and #6 had litters with one affected male each, #60 and #3, respectively. Although DNA samples from animals #48, #51, #53 and #60 were not available, blood parameters and medical reports about recurrent hemorrhagic episodes were provided (*Online*

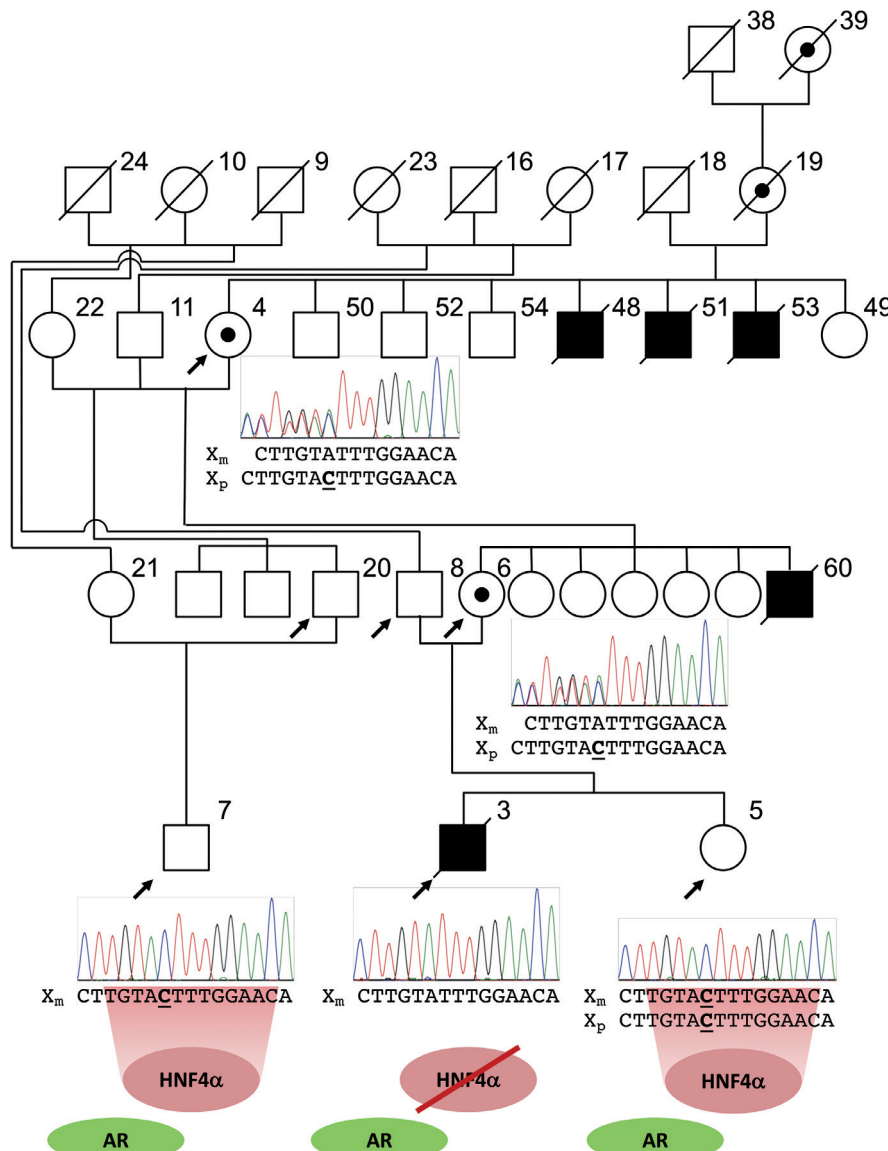


Figure 1. Pedigree section of the hemophilia B Leyden Hovawart family and DNA sequence comparison of the mutant hepatocyte nuclear factor 4 α /androgen receptor binding site in the promoter of canine *F9* in the hemophilic male (#3) and relatives (#4 grandmother, #5 sister, #6 mother, #7 cousin). Pedigree symbols are according to the standardized human pedigree nomenclature.⁶¹ Individuals are pseudonymized using internal identities. DNA samples were available from individuals indicated with an arrow. DNA sequences of heterozygous bitches #4 and #6 (female carriers) show overlapping peaks with similar heights 5' of the deletion position. For males #48, #51, #53 and #60, signs of hemophilia (*Online Supplementary Table S1*) were reported and the dogs had to be euthanized after recurrent hemorrhages. X_m : maternal X chromosome; X_p : paternal X chromosome; HNF4 α : hepatocyte nuclear factor 4 α binding site (consensus sequence: 5'-TGNACTTG-3')^{21,48}; AR: 3'-part of the androgen receptor binding site (consensus sequence: 5'-AGNACANNNTGTNCT-3')^{21,48}.

Supplementary Table S1). These males had increased aPTT ratios of 2.93 (#53) to 4.76 (#51) indicative of defects in the intrinsic coagulation pathway and also reduced FIX concentrations in the blood as is normally the case in hemophilia B. The affected dog #3 presented only 2% of the standard FIX concentration. The female carriers #4 and #6 showed aPTT ratios within the reference range. FIX concentrations, however, were slightly below the reference range (#6). This was not surprising as it has been shown that minute reductions in FIX concentrations might not always be reflected in an aPTT increase because of the sensitivity of commercial reagents.⁴³ The clinical signs together with the blood coagulation parameters and X-linked transmission supported the diagnosis of hemophilia B. The definite clinical diagnosis prompted us to search for the molecular cause initially on the DNA level. The canine *F9* gene is located on chromosome X (CFAX) between positions 109,501,341 (transcription start site) and 109,533,798 and has a length of 32,458 bp (NC_006621.3, CanFam3.1). The canine *F9* gene, similarly to that of other mammals, has eight exons with an open reading frame of 1,356 bp coding for 452 amino acids.⁴⁴ DNA of female carriers #4 and #6 was subjected to whole genome sequencing and aligned to the canine reference *F9* gene sequence. Surprisingly, only six sequence variants outside the coding regions of *F9* were identified (Table 1). Five variants were located in introns and were excluded as the cause of the hemophilia B in the Hovawarts because these variants were also detected in unaffected controls. The remaining variant (deletion) was located in the promoter of *F9* 73 bp upstream of the start codon (Online Supplementary Figures S1 and S3). As this deletion was located within a putative transcription factor binding site of HNF4α and AR, which had been shown in humans to be important for *F9* expression and mutated in hemophilia B Leyden and Brandenburg,^{31,32} this position was analyzed in more detail.

Figure 1 shows the segregation of the nucleotide deletion in the affected Hovawart family. The female carriers #4 and #6 were heterozygous, as evidenced by the overlap-

ping peaks with similar heights 5' of the deletion position. The affected male #3 was hemizygous for the deleted allele whereas his sister #5 and cousin #7 were homozygous wildtype. Genotyping of 1,298 dogs (including 83 different breeds, 720 unrelated Hovawarts, and 12 Hovawart family members) demonstrated the occurrence of the deletion only among members of the affected Hovawart family (Table 2, Online Supplementary Table S2). To provide proof that the deletion represented the causative genetic variant and resulted in the low expression of *F9*, functional analyses using electrophoretic mobility shift and luciferase reporter assays were performed.

As shown in Figure 2, no binding of recombinant HNF4α to the mutated promoter region was detected. On the other hand, the AR lysate clearly showed binding to both fragments and hence the deletion seems not to influence AR binding to the androgen-responsive element in the canine *F9* promoter. To analyze the effect of the promoter variant on *F9* expression, wildtype and mutated promoter fragment luciferase constructs were transfected into Hep G2 cells. As shown in Figure 3 the mutated promoter fragment resulted in a statistically highly significant ($P=2.2 \times 10^{-9}$) reduction of the relative response ratio to approximately 34.6% of the wildtype promoter in the presence of C/EBP (+ C/EBP). C/EBP is clearly also an important transcription factor in the regulation of the canine *F9* promoter as shown when C/EBP was not co-transfected (- C/EBP). In the absence of C/EBP the relative response ratio of the wildtype promoter was 29.8%. On the other hand there were no significant differences between the mutated promoter fragment (+/- C/EBP) and the wildtype promoter fragment (- C/EBP). As for variants of the HNF4α site, disruption of the C/EBP binding site has also been shown to be causative for hemophilia B in humans.^{45,46}

Discussion

As in humans, hemophilia A and B are also rare diseases in dogs caused by sequence variants in the coagulation

Table 1. DNA sequence variants in the canine *F9* gene determined by next-generation sequencing of DNA of animals #4 and #6.

Position	Ref/Alt ^a	Gene region	HGVs ^b g.
X:109501492	C/-	5'-flanking region	NC_006621.3:g.109501492delC
X:109504229	C/-	intron 1	NC_006621.3:g.109504229delC
X:109505462	-/AG	intron 1	NC_006621.3:109505462_109505463insAG
X:109507446	-/A	intron 2	NC_006621.3:109507446_109507446insA
X:109510986	G/A	intron 3	NC_006621.3:g.109510986G>A
X:109524055	A/G	intron 6	NC_006621.3:g.109524055A>G

^aRef/Alt: reference nucleotide/alternate nucleotide; ^bHGVs: Human Genome Variation Society (<http://www.hgvs.org>).

Table 2. *F9* genotype frequencies

Genotype	HB ^a affected (n=1)	HB carrier (n=2)	Hovawart Control, related (n=12)	Control, unknown relationship (n= 720)	Other breeds ^b Controls (n= 567)
C/C			12	720	567
C/-		2			
-/-	1				

^aHB: hemophilia B; ^bdog breeds used as controls are listed in Online Supplementary Table S2.

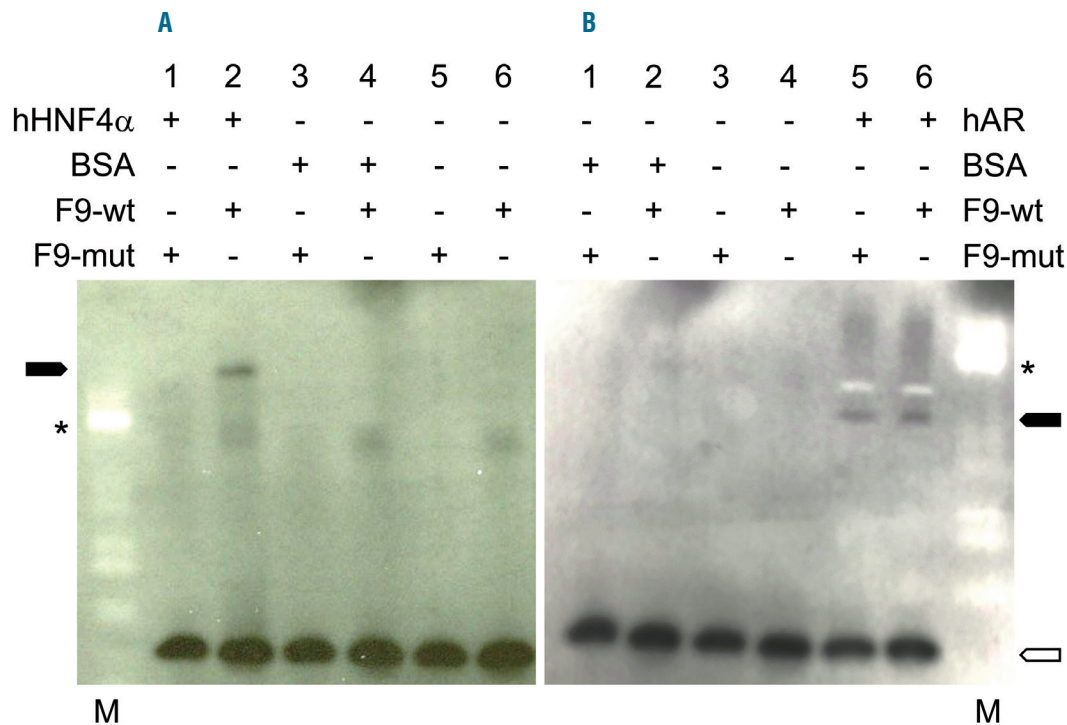


Figure 2. Analysis of hepatocyte nuclear factor 4α and androgen receptor binding of wildtype and mutated *F9* promoter regions using an electrophoretic mobility shift assay. (A, B) Human hepatocyte nuclear factor 4α (HNF4α) and (B) androgen receptor (AR) were used to bind biotin-labeled wildtype and mutated *F9* promoter fragments (F9-wt, F9-mut). Specific shifted bands (solid arrowheads) are detected in lane 2 (A) for HNF4α and lanes 5 and 6 (B) for AR. To test specificity, binding reactions were also performed using bovine serum albumin [BSA; lanes 3 and 4 (A), lanes 1 and 2 (B)]. In lanes 5 and 6 (A) and lanes 3 and 4 (B) no protein was added. Binding reactions were separated on 12% Tris-Glycine gels. X-ray films were cropped using GIMP 2.8.22. The 70 kDa protein marker band (PageRuler Prestained Protein Ladder, Fermentas) is indicated with an asterisk (lane M). The open arrowhead indicates unbound, free DNA.

factor VIII (*F8*) and IX (*F9*) gene, respectively.⁴⁷ Since the 1960s cases of canine hemophilia B have been reported and in 1989 the first description of the molecular cause in a Cairn Terrier population at the Francis Owen Blood Research Laboratory was published.^{5,8,9,44}

Since then only six further types of variants, all of them affecting the coding region of the *F9* gene, *i.e.* deletions, insertions, missense variants, have been described as causative for canine hemophilia B.^{10-12,15,16} The identification of a causative promoter variant in the Hovawart dogs described here is therefore unique in two respects: it is the first regulatory variant described in dogs and secondly this variant resembles a specific subtype of hemophilia B, known as hemophilia B Leyden, in humans.²¹ Hitherto, 21 distinct variants in the human *F9* promoter have been determined in families affected by hemophilia B Leyden.⁴⁸ These variants cluster in the so-called Leyden-specific region (LSR) and interfere with the binding of different transcription factors, *e.g.* AR, HNF4α, ONECUT, and C/EBPα.⁴⁹ The deletion identified in the Hovawart dogs was located 73 bp upstream of the start codon of the canine *F9* gene corresponding to position -23 in the third human promoter cluster harboring the overlapping binding sites of AR and HNF4α.⁴⁸ Similar to analyses in humans, it was possible to show by electrophoretic mobility shift assay that the deletion in the canine promoter abolished HNF4α binding because it affects the highly conserved core sequence of the HNF4α binding motif.²⁴ On the other hand, binding of AR was not affected. This might be due to the fact that AR DNA-binding

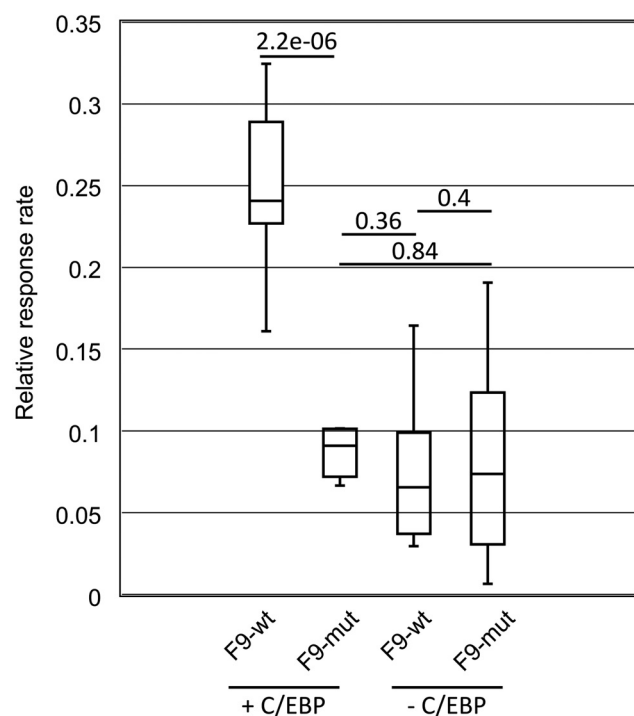


Figure 3. Dual-luciferase reporter analysis of *F9* promoter activities in Hep G2 cells. Box and whisker plot showing the change of relative response ratios (RRR) between the wildtype (F9-wt) and mutant promoter (F9-mut) gene constructs. The lines in the boxes represent the medians. Whiskers indicate minimum and maximum RRR values. Values were normalized as described in the Online Supplementary Methods. P-values are indicated.

sites display an exceptional amount of sequence variation.⁵⁰ Although the C-deletion is located in the consensus TGTNCT-motif of class I AR-binding sites several alternative motives, e.g. TGTTC in the stomatin-like protein 3 gene or TGTATC in the prostate-specific antigen gene enhancer III region, have been reported.⁵⁰⁻⁵² Therefore, it can be assumed that the affected males would have recovered from hemophilia during puberty. To what extent a sex- or possibly also age-dependent amelioration of the hemophilia B Leyden, as described in humans and also a mouse model, might be present in the affected Hovawarts remains to be determined.^{49,53} In this respect the Hovawart family could also be of interest as a model in the comparative analysis of age-dependent normalization of *F9* expression in symptomatic carriers of hemophilia B Leyden in humans.

To analyze whether the deletion not only affected HNF4 α binding but also resulted in a downregulation of *F9* expression, we performed luciferase assays. Although relative response ratios do not directly reflect *F9* mRNA or FIX levels *in vivo*, silencing of the promoter activity due to the deletion was obvious from these experiments. Similar findings have been made using HNF4 α -null mice.²⁶ In the murine model it was shown that expression of factors V, XI, XII, and XIIIb directly depends on hepatic HNF4 α . Northern blot analysis also demonstrated that *F9* expression was decreased with a significantly prolonged aPTT in the HNF4 α -null mice.²⁶ The finding that *F9* expression was not completely dependent on HNF4 α is in agreement with the observation that control of *F9* transcription in mice and humans is complex and depends on a plethora of factors.²⁶ For instance, in earlier studies it was shown, using DNaseI footprinting, that there are further binding

sites of HNF4 α and other factors, e.g. ARP1, COUP/Ear3, in the human *F9* promoter influencing *F9* expression.⁵⁴ However, the classical HNF4 α binding site at position -26 to -19 only binds HNF4 α . When analyzing the canine *F9* promoter using transcription factor binding site prediction algorithms, further potential binding motifs for HNF4 α and other transcription factors were predicted (*data not shown*).⁵⁵⁻⁵⁷ As ectopic expression of *F9* *in vivo* can be excluded or at least ignored according to recent RNA-sequencing analyses,⁵⁸ a remaining reduced hepatic activity of the mutated promoter is in agreement with the clinical findings of residual FIX activity in the affected males (*Online Supplementary Table S4*) and the results of the electrophoretic mobility shift assay showing binding of AR in androgen-dependent promoter activation.

In summary, we have identified and elucidated the causative genetic variant for hemophilia B Leyden in Hovawarts. This is the first report on a single nucleotide deletion within the binding sites of HNF4 α and AR in the *F9* promoter causing hemophilia B Leyden in dogs. As the deletion only abolishes the binding of HNF4 α , it can be assumed that male dogs will most likely recover during puberty, as reported in humans.^{30,59,60} However, to prevent any risk of a further propagation of the disorder, genotyping of females is recommended in further breeding.

Acknowledgments

The authors are grateful to S. Pach for expert technical assistance and L. Binder for support. The owners of Hovawarts who have provided blood samples are thanked for their generous help. A. Leutz and E. Kowenz-Leutz are thanked for providing the C/EBP α expression vector. S. Shan and F. Xu are supported by a fellowship from the Chinese Scholarship Council (CSC).

References

- Dolan G, Benson G, Duffy A, et al. Haemophilia B: where are we now and what does the future hold? *Blood Rev.* 2018;32(1):52-60.
- Green P. The 'royal disease'. *J Thromb Haemost.* 2010;8(10):2214-2215.
- Rogaev EI, Grigorenko AP, Faskhutdinova G, Kittler EL, Moliaka YK. Genotype analysis identifies the cause of the "royal disease". *Science.* 2009;326(5954):817.
- Rallapalli PM, Kembal-Cook G, Tuddenham EG, Gomez K, Perkins SJ. An interactive mutation database for human coagulation factor IX provides novel insights into the phenotypes and genetics of hemophilia B. *J Thromb Haemost.* 2013;11(7):1329-1340.
- Mustard JF, Rowsell HC, Robinson GA, Hoeksema TD, Downie HG. Canine haemophilia B (Christmas disease). *Br J Haematol.* 1960;6:259-266.
- Mustard JF, Basser W, Hedgard G, Secord D, Rowsell HC, Downie HG. A comparison of the effect of serum and plasma transfusions on the clotting defect in canine haemophilia B. *Br J Haematol.* 1962;8:36-42.
- Parks BJ, Brinkhous KM, Harris PF, Penick GD. Laboratory detection of female carriers of canine hemophilia. *Thromb Diath Haemorrh.* 1964;12:368-376.
- Rowsell HC, Downie HG, Mustard JF, Leeson JE, Archibald JA. A disorder resembling hemophilia B (Christmas disease) in dogs. *J Am Vet Med Assoc.* 1960;137:247-250.
- Evans JP, Brinkhous KM, Brayer GD, Reisner HM, High KA. Canine hemophilia B resulting from a point mutation with unusual consequences. *Proc Natl Acad Sci U S A.* 1989;86(24):10095-10099.
- Brooks MB, Gu W, Ray K. Complete deletion of factor IX gene and inhibition of factor IX activity in a Labrador retriever with hemophilia B. *J Am Vet Med Assoc.* 1997;211(11):1418-1421.
- Gu W, Brooks M, Catalfamo J, Ray J, Ray K. Two distinct mutations cause severe hemophilia B in two unrelated canine pedigrees. *Thromb Haemost.* 1999;82(4):1270-1275.
- Brooks MB, Gu W, Barnas JL, Ray J, Ray K. A Line 1 insertion in the factor IX gene segregates with mild hemophilia B in dogs. *Mamm Genome.* 2003;14(11):788-795.
- Feldman DG, Brooks MB, Dodds WJ. Hemophilia B (factor IX deficiency) in a family of German shepherd dogs. *J Am Vet Med Assoc.* 1995;206(12):1901-1905.
- Kooistra HS, Slappendel RJ. [A young male mongrel with hemophilia-B (Christmas disease)]. *Tijdschr Diergeneesk.* 1991;116(6):281-285.
- Mausier AE, Whitlark J, Whitney KM, Lothrop CD Jr. A deletion mutation causes hemophilia B in Lhasa Apso dogs. *Blood.* 1996;88(9):3451-3455.
- Mischke R, Kuhnlein P, Kehl A, et al. G244E in the canine factor IX gene leads to severe haemophilia B in Rhodesian Ridgebacks. *Vet J.* 2011;187(1):113-118.
- Nakata M, Sakai M, Sakai T. Hemophilia B in a crossbred Maltese dog. *J Vet Med Sci.* 2006;68(11):1223-1224.
- Reitsma PH, Mandalaki T, Kasper CK, Bertina RM, Briet E. Two novel point mutations correlate with an altered developmental expression of blood coagulation factor IX (hemophilia B Leyden phenotype). *Blood.* 1989;73(3):743-746.
- Reitsma PH, Bertina RM, Ploos van Amstel JK, Riemens A, Briet E. The putative factor IX gene promoter in hemophilia B Leyden. *Blood.* 1988;72(3):1074-1076.
- Veltkamp JJ, Meilof J, Remmelts HG, van der Vlerk D, Loeliger EA. Another genetic variant of haemophilia B: haemophilia B Leyden. *Scand J Haematol.* 1970;7(2):82-90.
- Funnell AP, Crossley M. Hemophilia B Leyden and once mysterious cis-regulatory mutations. *Trends Genet.* 2014;30(1):18-23.
- Picketts DJ, Lillicrap DP, Mueller CR. Synergy between transcription factors DBP and C/EBP compensates for a haemophilia B Leyden factor IX mutation. *Nat Genet.* 1993;3(2):175-179.
- Yuan X, Ta TC, Lin M, et al. Identification of an endogenous ligand bound to a native orphan nuclear receptor. *PLoS One.* 2009;4(5):e5609.
- Bolotin E, Liao H, Ta TC, et al. Integrated approach for the identification of human hepatocyte nuclear factor 4 α target genes using protein binding microarrays.

- Hepatology. 2010;51(2):642-653.
25. Weltmeier F, Borlak J. A high resolution genome-wide scan of HNF4alpha recognition sites infers a regulatory gene network in colon cancer. *PLoS One*. 2011;6(7):e21667.
 26. Inoue Y, Peters LL, Yim SH, Inoue J, Gonzalez FJ. Role of hepatocyte nuclear factor 4alpha in control of blood coagulation factor gene expression. *J Mol Med (Berl)*. 2006;84(4):334-344.
 27. Belvini D, Salviato R, Radossi P, et al. Molecular genotyping of the Italian cohort of patients with hemophilia B. *Haematologica*. 2005;90(5):635-642.
 28. Ghanem N, Costes B, Martin J, et al. Twenty-four novel hemophilia B mutations revealed by rapid scanning of the whole factor IX gene in a French population sample. *Eur J Hum Genet*. 1993;1(2):144-155.
 29. Ketterling RP, Liu JZ, Liao D, et al. Two novel factor IX promoter mutations: incremental progress towards 'saturation in vivo mutagenesis' of a human promoter region. *Hum Mol Genet*. 1995;4(4):769-770.
 30. Reijnen MJ, Peerlinck K, Maasdam D, Bertina RM, Reitsma PH. Hemophilia B Leyden: substitution of thymine for guanine at position -21 results in a disruption of a hepatocyte nuclear factor 4 binding site in the factor IX promoter. *Blood*. 1993;82(1):151-158.
 31. Heit JA, Ketterling RP, Zapata RE, Ordonez SM, Kasper CK, Sommer SS. Hemophilia B Brandenberg-type promoter mutation. *Haemophilia*. 1999;5(1):73-75.
 32. Crossley M, Ludwig M, Stowell KM, De Vos P, Olek K, Brownlee GG. Recovery from hemophilia B Leyden: an androgen-responsive element in the factor IX promoter. *Science*. 1992;257(5068):377-379.
 33. Pfeiffer I, Volkel I, Taubert H, Brenig B. Forensic DNA-typing of dog hair: DNA-extraction and PCR amplification. *Forensic Sci Int*. 2004;141(2-3):149-151.
 34. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988;16(3):1215.
 35. Brown J, Pirrung M, McCue LA. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*. 2017 Jun 9. [Epub ahead of print]
 36. Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*. 2005;438(7069):803-819.
 37. Burland TG. DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol*. 2000;132:71-91.
 38. Clewley JP. Macintosh sequence analysis software. DNASTAR's LaserGene. *Mol Biotechnol*. 1995;3(3):221-224.
 39. Jiang M, Li H, Zhang Y, et al. Transitional basal cells at the squamous-columnar junction generate Barrett's oesophagus. *Nature*. 2017;550(7677):529-533.
 40. Lawson B, Robinson RA, Fernandez JR, et al. Spatio-temporal dynamics and aetiology of proliferative leg skin lesions in wild British finches. *Sci Rep*. 2018;8(1):14670.
 41. Eggers C, Hook B, Lewis S, Strayer C, Landreman A. Designing a bioluminescent reporter assay: normalization. 2016 [cited 14.11.2018]; Available from: <http://www.promega.de/resources/pub-hub/designing-a-bioluminescent-reporter-assay-normalization/>
 42. Dögel M. Working Dog. 2019 [cited 01.02.2019]; Available from: <https://en.working-dog.com>
 43. Mischke R. [Comparison of factor VIII:C and factor IX sensitivity of different commercial APTT reagents for canine plasma]. *Berl Munch Tierarztl Wochenschr*. 1999;112(10-11):394-399.
 44. Evans JP, Watzke HH, Ware JL, Stafford DW, High KA. Molecular cloning of a cDNA encoding canine factor IX. *Blood*. 1989;74(1):207-212.
 45. Bentley AK, Rees DJ, Rizza C, Brownlee GG. Defective propeptide processing of blood clotting factor IX caused by mutation of arginine to glutamine at position -4. *Cell*. 1986;45(3):343-348.
 46. Crossley M, Brownlee GG. Disruption of a C/EBP binding site in the factor IX promoter is associated with hemophilia B. *Nature*. 1990;345(6274):444-446.
 47. Mischke R. [Hemophilia A and B in dogs]. *Tierarztl Prax Ausg K Kleintiere Heimtiere*. 2012;40(1):44-53; quiz 54.
 48. Funnell AP, Wilson MD, Ballester B, et al. A CpG mutational hotspot in a ONECUT binding site accounts for the prevalent variant of hemophilia B Leyden. *Am J Hum Genet*. 2013;92(3):460-467.
 49. Kurachi S, Huo JS, Ameri A, Zhang K, Yoshizawa AC, Kurachi K. An age-related homeostasis mechanism is essential for spontaneous amelioration of hemophilia B Leyden. *Proc Natl Acad Sci U S A*. 2009;106(19):7921-7926.
 50. Reid KJ, Hendy SC, Saito J, Sorensen P, Nelson CC. Two classes of androgen receptor elements mediate cooperativity through allosteric interactions. *J Biol Chem*. 2001;276(4):2943-2952.
 51. Wilson S, Qi J, Filipp FV. Refinement of the androgen response element based on ChIP-Seq in androgen-insensitive and androgen-responsive prostate cancer cell lines. *Sci Rep*. 2016;6:32611.
 52. Schuur ER, Henderson GA, Kmetec LA, Miller JD, Lamparski HG, Henderson DR. Prostate-specific antigen expression is regulated by an upstream enhancer. *J Biol Chem*. 1996;271(12):7043-7051.
 53. Hildyard C, Keeling D. Effect of age on factor IX levels in symptomatic carriers of hemophilia B Leyden. *Br J Haematol*. 2015;169(3):448-449.
 54. Naka H, Brownlee GG. Transcriptional regulation of the human factor IX promoter by the orphan receptor superfamily factor, HNF4, ARP1 and COUP/Ear3. *Br J Haematol*. 1996;92(1):231-240.
 55. Fang B, Mane-Padros D, Bolotin E, Jiang T, Sladek FM. Identification of a binding motif specific to HNF4 by comparative analysis of multiple nuclear receptors. *Nucleic Acids Res*. 2012;40(12):5343-5356.
 56. Farre D, Roset R, Huerta M, et al. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res*. 2003;31(13):3651-3653.
 57. Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics*. 2002;18(2):333-334.
 58. Fagerberg L, Hallstrom BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. 2014;13(2):397-406.
 59. Reijnen MJ, Sladek FM, Bertina RM, Reitsma PH. Disruption of a binding site for hepatocyte nuclear factor 4 results in hemophilia B Leyden. *Proc Natl Acad Sci U S A*. 1992;89(14):6300-6303.
 60. Morgan GE, Rowley G, Green PM, Chisholm M, Giannelli F, Brownlee GG. Further evidence for the importance of an androgen response element in the factor IX promoter. *Br J Haematol*. 1997;98(1):79-85.
 61. Bennett RL, French KS, Resta RG, Doyle DL. Standardized human pedigree nomenclature: update and assessment of the recommendations of the National Society of Genetic Counselors. *J Genet Couns*. 2008;17(5):424-433.

Christmas disease in a Hovawart family resembling human hemophilia B Leyden is caused by a single nucleotide deletion in a highly conserved transcription factor binding site of the *F9* gene promoter

Bertram Brenig,^{1*} Lilith Steingraber^{1*}, Shuwen Shan,¹ Fangzheng Xu,¹ Marc Hirschfeld,^{1,2} Reiner Andag,³ Mirjam Spengler,⁴ Elisabeth Dietschi,⁴ Reinhard Mischke⁵ and Tosso Leeb⁴

¹University of Göttingen, Institute of Veterinary Medicine, Göttingen, Germany; ²Department of Obstetrics and Gynecology, Freiburg University Medical Center, Freiburg, Germany; ³University Medical Center Göttingen, Institute for Clinical Chemistry, Göttingen, Germany; ⁴Institute of Genetics, University of Bern, Bern, Switzerland and ⁵Small Animal Clinic, University of Veterinary Medicine Hannover Foundation, Hannover, Germany

**BB and LS contributed equally to this work.*

©2019 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2018.215426

Received: December 26, 2018

Accepted: March 6, 2019.

Pre-published: March 7, 2019.

Correspondence: *BERTRAM BRENIG* - bbrenig@gwdg.de

Supplementary Information

Supplementary Methods

Coagulation assays and FIX activity measurement

APTT (activated partial thromboplastin time) was measured coagulometrically using different commercially available activating reagents according to the manufacturer's test instructions. To standardize measurement results performed in different laboratories, ratio values (aPTT patient/median aPTT of healthy dogs) were calculated and reported. FIX activity was measured coagulometrically using human FIX deficient plasma and a commercial human aPTT reagent for activation. Canine pooled plasmas were used as reference (activity defined as 100 %).

Next generation sequencing (NGS) and genotyping

A 450 bp library was prepared from genomic DNA with the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs GmbH, Frankfurt, Germany) following the manufacturer's instructions. Library quality was evaluated with Agilent2100 Bioanalyzer. The following assembly options were used: mer-size 31 nt, min. match percentage 98, high layout stringency, min. aligned length 120 nt, min. layout length 50 nt, max. gap size 5 nt. Duplicate reads were combined and clonal reads removed. For #4 4,590 and for #6 4,982 consistent paired reads were assembled. The sample wise insert size metrics for high quality aligned reads was median pair distance 383.1 bp (SD 117.46 bp, min. distance 151 bp, max. distance 894 bp) for #4 and 383.9 bp (SD 103.55 bp, min. distance 153 bp, max. distance 882 bp) for #6.

Electrophoretic mobility shift assay (EMSA)

Binding reactions included 2 µL 10 x binding buffer (100 mM Tris, 10 mM EDTA, 1 M KCl, 60% v/v Glycerol (86% solution), 0.1 mg/ml BSA, 0.5% Triton X-100, 1mM DTT; pH 7.5), 2 µg poly(dI-dC) and 1.2 µg human HNF4α or 4 µg poly(dI-dC) and 5 µg human AR lysate. As negative controls 1 pmol duplex DNA oligos were incubated without protein or with 1 µg BSA. Binding reactions were pre-incubated for 20 min on ice followed by 1 hour at room temperature after adding 1 pmol biotin-end labelled double-stranded oligonucleotide probes. The mixtures were loaded onto 12% Tris-Glycine gels (Invitrogen, USA). After electrophoresis at 80 V for 90 min (HNF4α) or 2 hours (AR), gels were blotted onto PVDF membranes (GE Healthcare Life Sciences, Germany) using a wet blotter for 30 min at 100 V. Membranes were crosslinked at 120 mJ/cm² using a commercial UV-light crosslinking instrument equipped with 254 nm bulbs for 1 minute.

Luciferase-Assay

Promoter fragment design was geared to an equivalence of the canine genomic situation choosing a respective distance between NC_006621.3:g.109,501,492delC and

the luciferase start codon. Recombinant pGL3 vectors were used for transformation of *E. coli* XL1-Blue according to the manufacturer's protocol. Plasmid DNA of 17 colonies of pGL3Basic+970bpinsertF9_MT and 37 colonies of pGL3Basic+971bpinsertF9_WT were isolated using Promega PureYield Plasmid Miniprep Kit (Promega, Mannheim, Germany) and sequenced for validation. A validated clone of each construct was incubated in LB-medium and plasmid DNA was isolated using Qiagen Plasmid Maxi Kit (Qiagen, Hilden, Germany).

For analysis of promoter activity human hepatoma derived cell line Hep G2 (ATCC HB-8065) was cultivated in Roti-CELL DMEM High Glucose (Carl Roth GmbH, Karlsruhe, Germany) ¹. Constructs were transfected using Effectene Transfection Reagent (Qiagen, Hilden, Germany). Firefly and *Renilla* luciferase luminescence was measured using the Dual-Glo Luciferase Assay System (Promega, Mannheim, Germany) on a Tecan GENios Pro 96/384 Multifunction Microplate Reader (Tecan GmbH, Crailsheim, Germany) with the analysis software XFlour v4.64 after cell lysis with Passive Lysis 5X Buffer (Promega, Mannheim, Germany). For normalization *Renilla* luciferase activity was measured by co-transfecting phRL-TK(Int⁻) (Promega, Mannheim, Germany). Experiments were repeated 5-times with two measurements each. Background luminescence values were subtracted from raw luminescence values. *Renilla* luciferase activities were used for normalization ².

Supplementary Figures

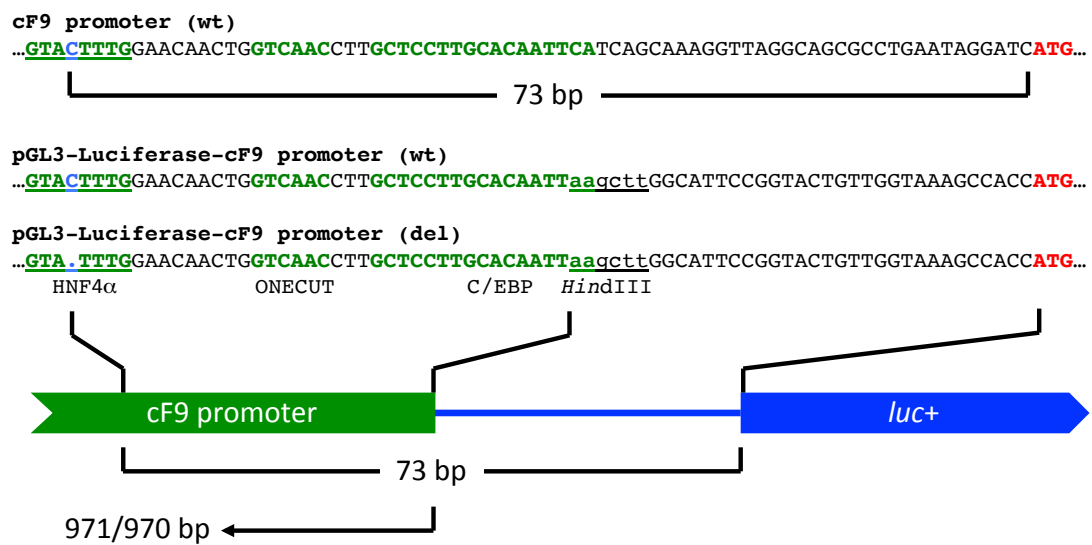


Figure S1. Structure of the pGL3 Luciferase Reporter Vector constructs

Fragments of the canine *F9* gene wild type (971 bp) and mutant (970 bp) promoter were amplified and cloned into the *HindIII* restriction site of the pGL3-Basic Vector multiple cloning site upstream of the *luc+* coding region. The distance of the C-deletion within the overlapping AR/HNF4 α binding sites 73 bp upstream the start codons (indicated in bold red) and the potential binding sites of ONECUT and C/EBP were cloned so that the distances of the positions in the promoter inserts to the luciferase reporter gene exactly mimic the *in vivo* genomic distances.

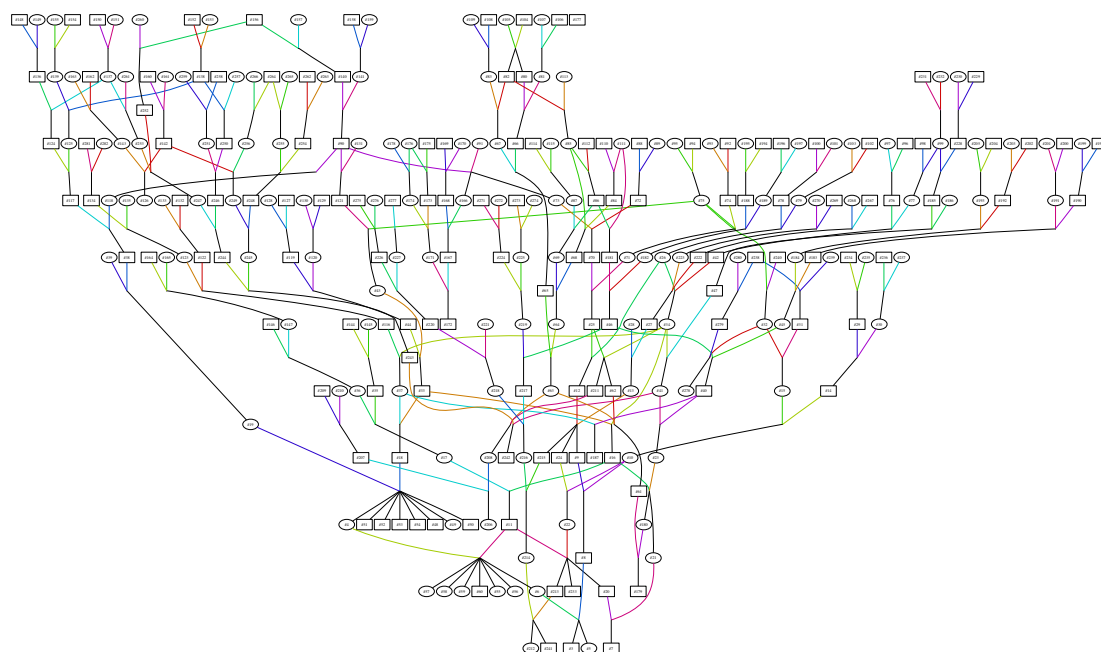


Figure S2. Detailed pedigree of the affected Hovawart family

The ancestry of the affected Hovawart #3 was reconstructed using the online dog breed database "Working dog" and pedigree data of individual dogs provided by the owners³. The pedigree was drawn using Pedigraph⁴. Males are indicated with squares, females with circles. Pseudonymized individual dog numbers are given within the pedigree symbols.

```

109501382
|
cF9 AAGATGAACA CTGCCTACAC TCTTAACCCA AGAGGCCACT GGAAATAGCC
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
hF9 GACCACTGCC CATTCTCTTC ACTTGTCCCA AGAGGCCATT GGAAATAGTC
| NF1-L
139530603

cF9 CAAAGACCCA CTGAGGGAGA TGGACACTGT TTCCCAGAAG TAAATACAGC
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
hF9 CAAAGACCCA TTGAGGGAGA TGGACATTAT TTCCCAGAAG TAAATACAGC

- /C (del)
|
cF9 TCAACTTGTA CTTTGGAACA ACTGGTCAAC CTTGCTCCTT GCACAATTCA
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
hF9 TCAGCTTGTA CTTTGGTACA ACTAATCGAC CTTACCACTT TCACAATCTG
AR HNF4a ONECUT C/EBP
. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
-26 -20 -6 +1 +13

cF9 TCAGCAAAGG TTAGGCAGCG CCTGAATAGG ATCATGGCAG AAGCATCGGG
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
hF9 CTAGCAAAGG TTATGCAGCG CGTGAACATG ATCATGGCAG AATCACCAGG
MetGlnAr gValAsnMet IleMetAlaG luSerProG1

cF9 CCTCGTCACC GTCTGCCTTT TAGGATATCT ACTCAGTGCC GAATGTGCAG
yLeuValThr ValCysLeuL euGlyTyrLe uLeuSerAla GluCysAla
hF9 CCTCATCACC ATCTGCCTTT TAGGATATCT ACTCAGTGCT GAATGTACAG
yLeuIleThr IleCysLeuL euGlyTyrLe uLeuSerAla GluCysThr

```

Figure S3. Sequence comparison of the human and canine *F9* promoter region including exon 1

Position 109501382 of the canine *F9* promoter (cF9) corresponds to accession number ENSCAFG00000018998 (CM000039.3, CanFam3.1) ⁵. Position 139530603 of the human *F9* promoter (hF9) corresponds to accession number NC_000023.11 (GRCh38.p12) ^{5, 6}. The start codons of exon 1 are indicated in red ⁵. Amino acids of the first exons are shown in 3-letter code below the corresponding DNA sequence. Identical nucleotide positions are depicted with colons. Consensus sequences of transcription factor binding sites identified in the human *F9* promoter are shown in bold and/or green: NF1-L (Nuclear factor-1 liver) ^{5, 7}, AR (Androgen receptor) ^{8, 9}, HNF4 α (Hepatocyte nuclear factor 4 α) ⁹⁻¹¹, ONECUT (ONECUT homeobox transcription factor 1 and 2) ^{5, 9, 11, 12}, C/EBP (CCAAT/enhancer-binding protein alpha) ^{5, 7, 9, 11}. In the overlapping binding sites of AR and HNF4 α , the binding site of HNF4 α is also underlined. The position (73 bp upstream of the ATG) of the C-deletion (-/C (del)) causative for Hemophilia B Leyden in the Hovawart family is indicated. The predicted HNF4 α binding site in the canine *F9* promoter using the online database PROMO 3.0 is shown in bold blue ^{13, 14}. The oligo

used for EMSA is underlined in cF9. The transcription start site of the human *F9* gene is indicated with +1^{5,7}. Bullet points and numbers below the human *F9* promoter DNA sequence mark nucleotide positions that have been shown to associated hemophilia B (+6: T>A⁹, +7: T>C⁹, +8: T>C⁹, +9: T>G⁹, +12: A>G⁹, +13: A>G, A>C or delA^{7,9}), hemophilia B Leyden (-5: A>T or A>G¹², -6: G>A or G>C¹², -19: G>C¹², -20: T>A or T>C^{8,11,12}, -21: T>G^{10,12}, -23: C>T or C>G¹², -24: A>G¹²) and hemophilia B Brandenburg (-26: G>C, G>T or G>A^{8,9,11,12}).

Table S1. Determination of hemophilia relevant blood parameters and medical reports

ID ^{a)}	Sex	aPTT ratio ^{b)}	FIX (%) ^{c)}	Medical report
#3	m	ND ^{d)}	2	severe bleeding after chipping, blood transfusion, death caused by blood loss
#4	f	1.08	92	
#5	f	1.1	64	
#6	f	1.1	58	
#7	m	1.03	83	
#48 ^{e)}	m	4.2	3	slight bleeding during second dentition, lameness/joint problems (age 4 months), several blood transfusions
#50	m	1.22	ND ^{d)}	
#51 ^{e)}	m	4.76	70	umbilical hernia with internal bleeding after surgery, blood transfusion, minor bleeding episodes
#52	m	3.19	110	
#53 ^{e)}	m	2.93	ND ^{d)}	recurrent slight bleeding, prolonged bleeding during second dentition, lameness/joint problems
#54	m	1.09	55	
#60 ^{e)}	m	4.47	5	severe bleeding after first vaccination (age 8 weeks)
Control 1	m	1.0	83	healthy unrelated Hovawart control
Control 2	f	0.98	>100	healthy unrelated Hovawart control

a) Animal IDs refer to Fig 1; b) aPTT ratio: reference range 0.88-1.14; c) FIX (%): FIX % of standard (reference range: 75-140%); d) ND: not determined; e) #48, #51, #53 and #60 had been euthanized.

Table S2. Dog breeds used for determination of *F9* genotype frequencies shown in Table 2

Airedale Terrier (n=1), Akita Inu (n=8), Alaskan Malamute (n=1), Appenzell Cattle Dog (n=8), Australian Cattle Dog (n=8), Australian Shepherd (n=8), Barbet (n=8), Barzoi (n=8), Bavarian Mountain Scent Hound (n=3), Beagle (n=8), Bearded Collie (n=8), Belgian Shepherd Dog (n=15), Bernese Mountain Dog (n=8), Border Collie (n=11), Boston Terrier (n=8), Boxer (n=2), Briard (n=8), Cairn Terrier (n=8), Canadian Sheepdog (n=1), Catalan Sheepdog (n=1), Chinese Crested Dog (n=8), Chihuahua (n=7), Cocker Spaniel (n=1), Dachshund (n=11), Dalmatian (n=1), German Hound (n=2), Doberman (n=8), Elo (n=8), Entlebuch Cattle Dog (n=8), Eurasier (n=1), Flat Coated Retriever (n=8), Fox Terrier (n=1), French Bulldog (n=9), German Shepherd (n=12), German Shorthaired Pointer (n=2), German Spaniel (n=10), Giant Schnauzer (n=6), Giant Spitz (n=8), Golden Retriever (n=15), Great Dane (n=8), Greyhound (n=8), Irish Terrier (n=8), Jack Russel Terrier (n=10), Keeshound (n=8), Kromfohrlander (n=8), Labrador Retriever (n=10), Lagotto Romagnolo (n=8), Landseer (n=8), Leonberger (n=8), Magyar Vizsla (n=1), Maltese (n=2), Xoloitzcuintle (n=6), Miniature Spitz (n=8), Miniature Pinscher (n=8), Miniature Poodle (n=2), Mudi (n=7), Mongrel (n=34), Newfoundland (n=7), Norwich Terrier (n=8), Nova Scotia Duck Tolling Retriever (n=8), Peruvian Hairless Dog (n=6), Polish Lowland Sheepdog (n=8), Pomeranian (n=8), Poodle (n=10), Portuguese Sheepdog (n=8), Pudelpointer (n=1), Pyrenean Sheepdog smoothfaced (n=2), Rhodesian Ridgeback (n=8), Saluki (n=4), Schapendoes (n=10), Scottish Terrier (n=8), Siberian Husky (n=9), Shi Tzu (n=1), Spanish Water Dog (n=8), St. Bernard (n=8), Tibet Terrier (n=8), Welsh Terrier (n=3), West Highland White Terrier (n=7), Whippet (n=8), Yorkshire Terrier (n=10)

Supplementary References

1. Knowles BB, Howe CC, Aden DP. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. *Science*. 1980;209(4455):497-499.
2. Jiwaji M, Daly R, Pansare K, et al. The Renilla luciferase gene as a reference gene for normalization of gene expression in transiently transfected cells. *BMC Mol Biol*. 2010;11(103).
3. Dögel M. Working Dog. 2019 [cited 01.02.2019]; Available from: <https://en.working-dog.com>
4. Garbe J, Da Y. Pedigraph: A Software Tool for the Graphing and Analysis of Large Complex Pedigree. User manual Version 2.4 ed. Department of Animal Science, University of Minnesota, 2008.
5. Pang CP, Crossley M, Kent G, Brownlee GG. Comparative sequence analysis of mammalian factor IX promoters. *Nucleic Acids Res*. 1990;18(22):6731-6732.
6. Reitsma PH, Bertina RM, Ploos van Amstel JK, Riemens A, Briet E. The putative factor IX gene promoter in hemophilia B Leyden. *Blood*. 1988;72(3):1074-1076.
7. Crossley M, Brownlee GG. Disruption of a C/EBP binding site in the factor IX promoter is associated with haemophilia B. *Nature*. 1990;345(6274):444-446.
8. Crossley M, Ludwig M, Stowell KM, De Vos P, Olek K, Brownlee GG. Recovery from hemophilia B Leyden: an androgen-responsive element in the factor IX promoter. *Science*. 1992;257(5068):377-379.
9. Funnell AP, Crossley M. Hemophilia B Leyden and once mysterious cis-regulatory mutations. *Trends Genet*. 2014;30(1):18-23.
10. Reijnen MJ, Peerlinck K, Maasdam D, Bertina RM, Reitsma PH. Hemophilia B Leyden: substitution of thymine for guanine at position -21 results in a disruption of a hepatocyte nuclear factor 4 binding site in the factor IX promoter. *Blood*. 1993;82(1):151-158.
11. Reijnen MJ, Sladek FM, Bertina RM, Reitsma PH. Disruption of a binding site for hepatocyte nuclear factor 4 results in hemophilia B Leyden. *Proc Natl Acad Sci U S A*. 1992;89(14):6300-6303.
12. Funnell AP, Wilson MD, Ballester B, et al. A CpG mutational hotspot in a ONECUT binding site accounts for the prevalent variant of hemophilia B Leyden. *Am J Hum Genet*. 2013;92(3):460-467.
13. Farre D, Roset R, Huerta M, et al. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res*. 2003;31(13):3651-3653.
14. Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics*. 2002;18(2):333-334.

CHAPTER 4

A Missense Mutation in the *KLF7* Gene Is a Potential Candidate Variant for Congenital Deafness in Australian Stumpy Tail Cattle Dogs

Fangzheng Xu[†], Shuwen Shan[†], Susan Sommerlad, Jennifer M. Seddon, Bertram Brenig^{*}

[†] These authors contributed equally to this work.

^{*} Corresponding author

The article was published online in *Genes*, 12(4), p.467. in March 2021. The full article can be found online at:

<https://doi.org/10.3390/genes12040467>

Author contributions:

Shuwen Shan performed data analysis including variant annotation, private variant filtering, gene conservation analysis, human deafness genes filter in three deaf dogs, and genotyping of candidate variant. She also wrote the original manuscript.

Article

A Missense Mutation in the *KLF7* Gene Is a Potential Candidate Variant for Congenital Deafness in Australian Stumpy Tail Cattle Dogs

Fangzheng Xu ^{1,†}, Shuwen Shan ^{1,†}, Susan Sommerlad ², Jennifer M. Seddon ² and Bertram Brenig ^{1,*}

¹ Institute of Veterinary Medicine, University of Goettingen, 37077 Göttingen, Germany; fangzheng.xu@stud.uni-goettingen.de (F.X.); shuwen.shan@stud.uni-goettingen.de (S.S.)

² School of Veterinary Science, The University of Queensland, Gatton, QLD 4343, Australia; s.sommerlad@uq.edu.au (S.S.); j.seddon1@uq.edu.au (J.M.S.)

* Correspondence: bbrenig@gwdg.de; Tel.: +49-551-3928383; Fax: +49-551-3933392

† These authors have equally contributed.

Abstract: Congenital deafness is prevalent among modern dog breeds, including Australian Stumpy Tail Cattle Dogs (ASCD). However, in ASCD, no causative gene has been identified so far. Therefore, we performed a genome-wide association study (GWAS) and whole genome sequencing (WGS) of affected and normal individuals. For GWAS, 3 bilateral deaf ASCDs, 43 herding dogs, and one unaffected ASCD were used, resulting in 13 significantly associated loci on 6 chromosomes, i.e., CFA3, 8, 17, 23, 28, and 37. CFA37 harbored a region with the most significant association ($-\log_{10}(9.54 \times 10^{-21}) = 20.02$) as well as 7 of the 13 associated loci. For whole genome sequencing, the same three affected ASCDs and one unaffected ASCD were used. The WGS data were compared with 722 canine controls and filtered for protein coding and non-synonymous variants, resulting in four missense variants present only in the affected dogs. Using effect prediction tools, two variants remained with predicted deleterious effects within the Heart development protein with EGF like domains 1 (*HEG1*) gene (NC_006615.3: g.28028412G>C; XP_022269716.1: p.His531Asp) and Kruppel-like factor 7 (*KLF7*) gene (NC_006619.3: g.15562684G>A; XP_022270984.1: p.Leu173Phe). Due to its function as a regulator in heart and vessel formation and cardiovascular development, *HEG1* was excluded as a candidate gene. On the other hand, *KLF7* plays a crucial role in the nervous system, is expressed in the otic placode, and is reported to be involved in inner ear development. 55 additional ASCD samples (28 deaf and 27 normal hearing dogs) were genotyped for the *KLF7* variant, and the variant remained significantly associated with deafness in ASCD ($p = 0.014$). Furthermore, 24 dogs with heterozygous or homozygous mutations were detected, including 18 deaf dogs. The penetrance was calculated to be 0.75, which is in agreement with previous reports. In conclusion, *KLF7* is a promising candidate gene causative for ASCD deafness.

Keywords: deafness; kruppel-like factor 7; genome wide association study; Australian stumpy tail cattle dog; brainstem auditory evoked response



Citation: Xu, F.; Shan, S.; Sommerlad, S.; Seddon, J.M.; Brenig, B. A Missense Mutation in the *KLF7* Gene Is a Potential Candidate Variant for Congenital Deafness in Australian Stumpy Tail Cattle Dogs. *Genes* **2021**, *12*, 467. <https://doi.org/10.3390/genes12040467>

Academic Editor: Jeffrey M. Kidd

Received: 19 November 2020

Accepted: 23 March 2021

Published: 24 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Deafness can cause several inconveniences for dogs (*Canis familiaris*, CFA), as more attention is required to avoid undetected danger. Deaf dogs are not suitable as working dogs because their training is more challenging than for normal hearing dogs. In addition, they are more likely to be startled and show more tendency to bite [1]. More than 100 modern dog breeds have been reported to be affected by congenital deafness [2]. Hence, deafness seems to be a common disorder among dogs, particularly in breeds such as the Dalmatian, Bull Terrier, English Setter, English Cocker Spaniel, and Australian Cattle Dog [3]. Hearing loss or deafness can be categorized mainly by five criteria in dogs: (1) Cause (genetic or nongenetic, inherited or acquired); (2) association with other diseases

or phenotypes (syndromic or non-syndromic); (3) number of affected ears (unilateral or bilateral); (4) degree of loss (partial or total); and (5) site of pathology (peripheral or central) [4]. Peripheral deafness can also be classified as inherited or acquired, congenital or late onset, and sensorineural or conductive. In dogs, three classifications of deafness are commonly seen, including inherited congenital sensorineural, acquired later-onset sensorineural, and acquired later-onset conductive deafness [5].

In dogs, congenital sensorineural deafness is common, resulting in total deafness in young puppies that is either unilateral or bilateral. Sensorineural deafness results from dysfunction of cochlea or spiral ganglion. While it can be a degenerative process that relates to aging, noise trauma, exposure to therapeutic drugs that have ototoxic side effects, and chronic conditions [6], it is frequently inherited and so linked to one or more genetic mutations. Some morphological studies in dogs showed congenital sensorineural deafness manifested hypoplasia or aplasia of the sensory cells in the organ of Corti, stria vascularis, macula sacculle, solidification, and calcification of tectorial membrane [7,8]. Congenital sensorineural deafness is usually, but not always, related to pigmentation genes in some breeds [3].

Diagnosis of canine deafness usually consists of behavioral or electrodiagnostic testing. The behavioral testing is often unreliable, especially for the unilateral deafness or partial hearing impairment cases. The response of dogs may be affected by psychology (e.g., anxiety or loss of interest) and other senses (e.g., visual cues, vibration, or even air movement) [9]. The brainstem auditory evoked response (BAER) is the averaged record of the electrical activity of the auditory pathway in response to externally applied acoustic stimuli [10]. Compared with behavioral testing, the BAER test is an objective diagnostic method, with the advantages of being easy to record, noninvasive, safe, short test time, and giving reliable results [11].

The Australian Stumpy Tail Cattle Dog (ASCD) is a unique breed with a natural bob-tail, which should be distinguished from the Australian Cattle Dog breed. ASCD is alert, watchful and obedient, and talented in working and controlling cattle. It has been recognized as a standardized breed since 1963 by the Australian National Kennel Council. For a long time, general opinion held that the origins of the Australian Stumpy Tail Cattle Dog arose from European herding dogs and the Australian Dingo. However, recently it has been suggested that the ancestors of the Australian Stumpy Tail Cattle Dog and the Australian Cattle Dog, sharing a common origin, arrived in Australia with early free settlers, as their unidentified companions, between 1788 and c. 1800 (Clark, Noreen R. A Dog for the Job. (in prep. 2020)). Each pup should undergo a BAER test because this breed has a high deafness prevalence (<https://www.akc.org/dog-breeds/australian-stump-tail-cattle-dog/> (accessed on 24 March 2021)). A research study of 315 ASCDs showed the incidence of congenital sensorineural deafness was 17.8% [12]. There was no evidence that congenital sensorineural deafness in ASCD has a left/right asymmetry or a sex-specific pattern, but there was a significant correlation between red (over blue) coat color and deafness [12].

No unique causative variants have been identified so far for any dog breeds, possibly in part due to the fact that deafness appears to be a comparatively heterogeneous disease as described above. In addition, there are several hypotheses about the inheritance pattern of congenital sensorineural deafness (reviewed by [1]). In Border Collies, for instance, Ubiquitin Specific Peptidase 31 (*USP31*) and RB Binding Protein 6 (*RBBP6*) have been associated with adult-onset deafness [13], whereas in the Doberman Pinscher, an insertion in Protein Tyrosine Phosphatase Receptor Type Q (*PTPRQ*) and a missense variant in Myosin VIIA (*MYO7A*) have been shown to be causative for a form of deafness that includes vestibular disease [14,15]. Although chromosome 2 (CFA2), 6, 14, 17, 27, and 29 have been associated with hearing loss in Dalmatians, no causative variants have been identified so far [16].

In ASCD, congenital sensorineural deafness has been linked to a chromosomal region on CFA10 [12]. However, within a potential candidate gene Sry-related Hmg-box gene 10 (*SOX10*) located in this region, no causative alterations were detected. A recent genome-

wide association study (GWAS) reported 14 chromosomes that were significantly associated with deafness in three canine breeds, and CFA3 was significantly associated with bilateral deafness in Australian Cattle Dogs [17]. In this study, three suggestive candidate genes near significantly associated regions were detected in these three dog breeds, including ATPase Na⁺/K⁺ Transporting Subunit Alpha 4 (*ATP1A4*), Transformation/Transcription Domain Associated Protein (*TRRAP*), and Potassium Inwardly Rectifying Channel Subfamily J Member 10 (*KCNJ10*) [17]. However, none have been convincingly identified as causative mutations.

To extend the identification of potential candidate genes causing deafness in ASCD we performed a genome-wide association study and whole genome sequencing (WGS) in deaf ASCD. We identified a unique missense variant in Kruppel-like factor 7 (*KLF7*) gene significantly associated with deafness in ASCDs. This variant was absent in 722 dogs of bioproject PRJN448733 (see below). As *KLF7* plays an important role in the nervous system, is expressed in the inner ear, and seems to be involved in inner ear development [18,19], it was a convincing candidate for ASCD deafness.

2. Materials and Methods

2.1. Ethical Statement

The collection of dog blood samples was done by S. Sommerlad at the time of BAER testing. The collection of samples was approved by the “Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit” (33.19-42502-05-15A506) according to §8a Abs. 1 Nr. 2 of the TierSchG. All ASCDs were tested and sampled under approval of The University of Queensland’s Animal Ethics Committee.

2.2. Phenotyping and Samples

Fifty-nine Australian Stumpy Tail Cattle Dogs (Table S1) from a previous study [12] were used in this study. BAER testing was performed on 59 dogs [20], 28 were normal hearing dogs and 31 were diagnosed as deaf, of which 10 were bilateral deaf, 12 were left-sided deaf, and 9 were right-sided deaf (Table S1). Three bilaterally deaf ASCDs (#217, #253 and #330), and one control dog with normal hearing (#326) were used for next generation sequencing. Dog #326 was a littermate of #330. These four dogs were female and red in color; all but #330 had a speckled coat. DNA was extracted using a salting-out method as described [12]. All samples were pseudonymized using internal IDs. Furthermore, data from two repository were used in this study. One repository contain Variant Call Format (VCF) data of 722 canine individuals (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA448733> (accessed on 24 March 2021)) [21]. It consists of 144 established breeds, 11 samples with mixed breed, 26 samples with unknown breed status, 104 village and feral dogs from different regions, and 54 wild canids from six species. An additional dataset consisted of 590 samples including 582 dogs from 126 breeds and 8 wolves (<https://www.ebi.ac.uk/ena/data/view/PRJEB32865> (accessed on 24 March 2021)) [22].

2.3. Next Generation Sequencing and Variant Calling

A total of 1.0 µg DNA per ASCD sample was used as input material for the DNA library preparations. Sequencing libraries were generated using NEBNext® DNA Library Prep Kit following manufacturer’s recommendations and indices were added to each sample. The genomic DNA was randomly fragmented to a size of 350bp by shearing, then DNA fragments were end polished, A-tailed, and ligated with the NEBNext adapter for Illumina sequencing, and further PCR enriched by P5 and indexed P7 oligos. The PCR products were purified (AMPure XP system) and resulting libraries were analyzed for size distribution by Agilent 2100 Bioanalyzer and quantified using real-time PCR. For #217, #253, #326, #330, a total of 599,770,692, 723,624,660, 743,641,356, 620,101,998 raw reads were obtained, respectively. Corresponding coverages were around 40× (paired-end reads, 2 × 150 bp).

Raw sequence data were aligned to dog genome CanFam3.1 using BWA 0.7.17 [23]. SAMtools 1.9 were used for format change and sorting of sequences [24]. Duplicates were marked by PICARD (<http://broadinstitute.github.io/picard/>) (accessed on 24 March 2021)). Variant calling was performed by GATK 4.1.3 with best practice pipeline [25].

2.4. Genome Wide Association Analysis

We used the VCF data obtained in the previous step for GWAS analysis. Three deaf dogs (#217, #253, #330) were used as cases. As ASCD is utilized for control and herding of cattle according to its breed standard (<http://www.fci.be/Nomenclature/Standards/351g01-en.pdf>) (accessed on 24 March 2021)), VCFs of 43 herding dogs from 15 breeds (Australian Cattle Dog, Bearded Collie, Belgian Malinois, Belgian Sheepdog, Belgian Ter-vuren, Berger Blanc Suisse, Berger Picard, Border Collie, Bouvier des Flandres, Entlebucher Sennenhund, Finnish Lapphund, German Shepherd Dog, Pembroke Welsh Corgi, Shetland Sheepdog, Spanish Water Dog) were extracted from the publicly available 722 canine VCF repository (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA448733>) (accessed on 24 March 2021)) [21]. Sample selection criteria were the same as described [21]. A total of 43 herding dogs and the normal hearing dog #326 were chosen as controls (Table S2). The VCF files of the 43 herding dogs and 4 ASCDs were merged by BCFtools 1.9 [24]. Filtering was done using VCFtools 0.1.13 with options `-max-alleles 2`, `-min-alleles 2`, `-min-meanDP 20`, `-minQ 20`, `-minGQ 20`, `-remove-indels`, `-max-missing 0.95`, `-maf 0.05`, `-hwe 0.001` [26]. After filtering, 857,343 variants remained and were further pruned by Linkage Disequilibrium with `-indep 1000 3 1` function in PLINK 1.90 [27]. The final data set consisted of 20,656 single nucleotide polymorphisms (SNPs). Principal component analysis (PCA) was performed using EIGENSOFT package [28]. GEMMA 0.98 was used for association analysis by case-control setting (3 deaf cases vs. one normal hearing ASCD and 43 herding dogs as controls) [29]. A univariate linear mixed model with sex, 5 principal components, and relatedness of 47 dog individuals for corrections was applied for the association test. Bonferroni threshold $-\log_{10}P$ ($0.01/20,656$) = 6.32 was utilized. Qqman package was used to generate Manhattan and quantile-quantile (QQ) plots [30]. The genomic inflation factor lambda was calculated with formula $\lambda = \text{median}(\text{qchisq}(1-p, 1))/\text{qchisq}(0.5, 1)$ where p is a vector of p values.

2.5. Next Generation Sequencing Data Analysis for Identification of Associated Variants

Data after variant calling were analyzed with SNP & Variation Suite 8.8.3 (Golden Helix Inc., Bozeman, MT, USA). SNPs and indels were set to missing with read depth ≤ 10 , genotype quality ≤ 15 , alt read ratios for Ref_Ref ≥ 0.15 , Ref_Alt outside 0.3 to 0.7, Alt_Alt ≤ 0.85 . Variants were analyzed using autosomal recessive and dominant models, respectively. In the autosomal recessive filtering model, 3 deaf ASCDs were set as Alt_Alt, control ASCD as Ref_Ref or Alt_Ref. In the autosomal dominant filtering model, the 3 deaf ASCDs were set as Alt_Alt or Alt_Ref and controls as Ref_Ref. To further narrow the range of candidate variants, we compared the common variants of deaf ASCDs with 722 canine genomes to identify private variants. The shared variants in the three deaf dogs were filtered by BCFtools 1.9 with 'isec' option. Private variants were annotated using SnpEFF software [31] to determine high (loss of function) and moderate (missense) impact variants (Ensembl transcripts release 101). These functional variants were further checked by Integrative Genome Viewer (IGV) software to obtain real high quality variants [32]. Variant effects were predicted by SIFT [33], PolyPhen-2 [34], and PROVEAN [35].

2.6. Genotyping of KLF7 Variant in Australian Stumpy Tail Cattle Dogs

Targeted genotyping of the KLF7 missense variant was performed in 59 ASCDs by PCR amplification using primers cfa_KLF7_Ex3_F (5'-AGACTCTCTCAGCCGTGGAT-3') and cfa_KLF7_Ex3_R (5'-GGCCAACCTGTACCACTACCT-3'), resulting in a 295 bp fragment. Genotyping of PCR products were implemented by RFLP analysis after cleavage with the restriction enzyme HinP1I (NEB). The wild type allele was cleaved into two

fragments, 236 bp and 59 bp, while the homozygous mutant remained uncut. Frequency distribution for alleles and genotypes was calculated using Fisher's Exact Test in these 59 ASCDs. Allelic and genotypic odds ratios were calculated according to [36].

2.7. Investigation of Human Deafness Genes in 3 Deaf Australian Stumpy Tail Cattle Dogs

Human hearing loss or deafness genes were queried using online software GLAD4U with "hearing loss" and/or "deafness" as keywords [37]. After combining the three query results, 346 genes were chosen for further analysis (Table S3). The variants of these gene regions (including 1000 bp up- and downstream regions) were extracted by BCFtools from VCF files of the three deaf ASCDs and annotated by SnpEFF software. Variants with high (loss of function) and moderate (missense) impacts were selected for further analysis (Ensembl transcripts release 101). The genotype information of the chosen variants was further checked in 722 canines.

3. Results

3.1. Genome Wide Association Analysis

The analysis was done using three bilateral deaf female dogs from three different litters. The hearing status of the individuals determined using BAER is shown in Tables 1 and S1.

Table 1. BAER (brainstem auditory evoked response) results of 4 Australian Stumpy Tail Cattle Dogs (ASCDs).

ID	Gender	Coat Colour	BAER Test Results
217	Female	Red speckled	Bilaterally Deaf
253	Female	Red speckled	Bilaterally Deaf
330	Female	Red	Bilaterally Deaf
326	Female	Red speckled	Normal Hearing

The three affected ASCDs were compared with 44 control dogs. 13 SNPs on 6 chromosomes (CFA3, 8, 17, 23, 28, 37) above the Bonferroni significance level were identified. The QQ-plot indicated that some associations might be due to population substructure. Associated SNPs are shown in Figure 1 and summarized in Table 2. The majority of the significantly associated SNPs (7/13) were located on CFA37 including SNP chr37:44793 (position according to CanFam3.1) with the highest $-\log_{10}p$ -value = 20.02. A search for large structural variants (SVs) flanking the significantly associated regions on CFA3, 8, 17, 23, 28, and 37 using IGV was unsuccessful.

Table 2. Significantly associated SNPs above Bonferroni significance threshold (6.32).

CFA	Position	p-Value	Nearby Genes	Distance (bp)
3	90,987,932	2.67×10^{-8}	LCORL	186,575
8	62,032,863	5.93×10^{-13}	DGLUCY	19,884
17	1,977,343	1.73×10^{-7}	EIPR1	0
17	9,456,133	2.34×10^{-15}	TRIB2	204,307
23	50,096,314	3.04×10^{-7}	KCNAB1	0
28	21,516	4.50×10^{-9}	PTPN20	42,882
37	13,393	2.04×10^{-7}	WDR75	144,007
37	44,793	9.54×10^{-21}	WDR75	112,607
37	80,438	1.36×10^{-9}	WDR75	76,962
37	16,399,127	2.66×10^{-8}	CRYGD	25,757
37	22,102,392	6.48×10^{-10}	ABCA12	34,340
37	22,579,983	2.93×10^{-7}	FN1	57,573
37	22,711,697	1.10×10^{-8}	FN1	189,287

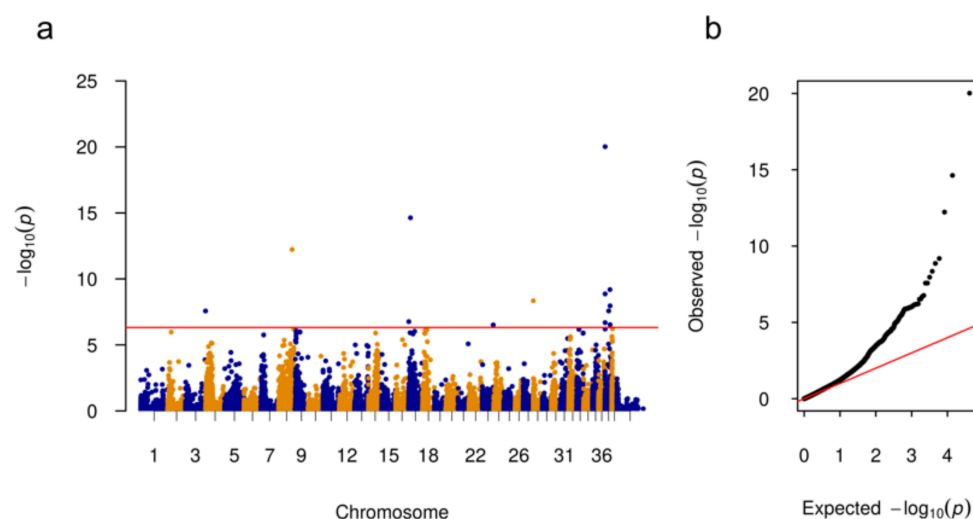


Figure 1. Manhattan and quantile–quantile (QQ) plots illustrating deafness associated chromosomal regions. **(a)** The Manhattan plot shows on the y -axis the negative log-base-10 of the p value for each of the polymorphisms in the genome (along the x -axis), when tested for differences in frequency between 3 bilateral deaf dogs (cases) and 44 controls ($1 \times$ ASCD, 43 herding dogs of 15 dog breeds). The red line indicates the Bonferroni significance threshold ($-\log_{10}(0.01/20,656) = 6.32$). **(b)** The QQ plot depicts the distribution of p -values of the genome-wide association study (GWAS) analysis and genomic inflation factor lambda is 1.20.

3.2. Whole Genome Sequencing Reveals Four Potential Variants

To further locate the candidate variants, next generation sequencing was performed in 3 deaf ASCDs (#217, #253, #330) and 1 normal hearing ASCD (#326). After quality control, a total of 4,208,002 SNPs and 2,298,760 indels were detected. According to previous deafness studies, sequence data were initially analyzed using a recessive model of inheritance. Using this model, 129,383 SNPs and 51,942 indels were detected. Using only variants that had been annotated and verified as mRNA transcripts (Ensembl release 101), 338 SNPs and 523 indels remained (Table S4). After filtering these variants against the 722 dog database, none of the homozygous Alt_Alt genotypes were exclusively present in the deaf ASCDs (Table S4). As there were no reports about such a high prevalence of deafness in the 722 control dogs and it can be assumed that the majority of the controls were hearing, these variants were presumably not causative.

As no associated variants were found using the recessive inheritance model, a dominant inheritance model was applied. In this analysis, private variants only present in the three deaf ASCDs (Alt_Alt and Alt_Ref) compared to the 722 controls (Ref_Ref) were filtered, resulting in 270,980 SNPs and 351,927 indels. After quality control and functional annotating, 167 protein-changing variants (58 SNPs and 109 indels) remained (Table S5). These variants were further filtered against #326 (normal hearing littermate of #330) assuming that this dog should be homozygous wild type under the supposed model. After this step, four missense variants remained as potential causative candidates (Table 3). Within the 722 control dogs, no homozygous Alt_Alt or heterozygous carriers were detected for these 4 missense variants. In an additional dataset consisting of 590 dog samples, only two heterozygous individuals (Brussels Griffon dogs) were determined for the Microtubule associated protein 6 (*MAP6*) gene variant. To deduce which of the variants could be causative for deafness, protein function prediction tools were used. As shown in Table 4 only the variants in Heart development protein with EGF like domains 1 (*HEG1*) and *KLF7* were predicted to be deleterious by at least two of the prediction tools.

Table 3. Genotype information of four potential causative variants for ASCD deafness.

Chr	HGVS Genome Position ^(a)	Variant Type	Gene ^(b)	#217	#253	#330	#326
13	NC_006595.3:g.60805542 C>T	missense variant	GC	C_T	C_T	C_T	C_C
21	NC_006603.3:g.23019999 C>T	missense variant	MAP6	C_T	C_T	C_T	C_C
33	NC_006615.3:g.28028412 G>C	missense variant	HEG1	G_C	G_C	G_C	G_G
37	NC_006619.3:g.15562684 G>A	missense variant	KLF7	A_A	A_A	A_G	G_G

^(a) Positions according to CanFam3.1; ^(b) GC: GC vitamin D binding protein, MAP6: Microtubule associated protein 6, HEG1: Heart development protein with EGF like domains 1, KLF7: Kruppel-like factor 7.

Table 4. Variant effect predicted by SIFT, PolyPhen-2, and PROVEAN.

Gene	Amino Acid Exchange	SIFT	Polyphen-2	PROVEAN
GC	p.Gly389Rrg	Tolerated	Benign	Neutral
MAP6	p.Arg486Cys	Affect protein function	Benign	Neutral
HEG1	p.His531Asp	Affect protein function	Unknown	Deleterious
KLF7	p.Leu173Phe	Affect protein function	Possibly damaging	Neutral

SIFT: <https://sift.bii.a-star.edu.sg> (accessed on 24 March 2021), Polyphen-2: <http://genetics.bwh.harvard.edu/pph2/index.shtml> (accessed on 24 March 2021), PROVEAN: <http://provean.jcvi.org/index.php> (accessed on 24 March 2021).

To further confirm the causative possibilities of the two remaining variants, their amino acid conservation was analyzed in the same 7 species. The missense variant in *HEG1* gene (NC_006615.3: g.28028412G>C) resulted in an amino acid exchange of p.His531Asp (XP_022269716.1). In *KLF7* gene (NC_006619.3: g.15562684G>A), the variant led to an exchange of p.Leu173Phe (XP_022270984.1). Especially in *KLF7*, the amino acid position seems to be highly conserved across several different species, as shown in Figure 2.

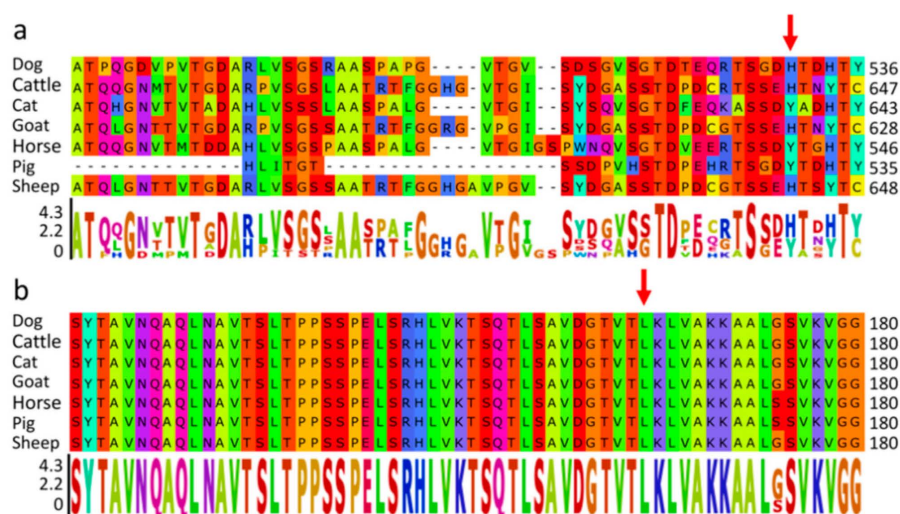


Figure 2. Cross-species comparison of variant amino acid positions in HEG1 and KLF7. Partial protein sequences of HEG1 (a) and KLF7 (b) flanking the variant amino acid positions were aligned using ClustalW (<https://www.ebi.ac.uk/Tools/msa/clustalo/> (accessed on 24 March 2021)). The variant positions are highlighted with a red arrow. Residual color scheme was referred from [38], sequence logos are shown according to [39].

3.3. Genotyping of KLF7 Variant in ASCDs

To verify the association of the *KLF7* variant with ASCD congenital deafness, 27 normal hearing and 28 deaf ASCDs (21 unilaterally and 7 bilaterally deaf dogs) were used to investigate the *KLF7* variant genotype distribution. As summarized in Table 5, 59 ASCDs including the 4 whole genome sequenced dogs were used to check the association of the *KLF7* missense variant with ASCD deafness. Four dogs were homozygous carriers (A_A) and 14 heterozygous (A_G) among the 31 deaf ASCDs. Within the 28 normal hearing AS-

CDs, 5 heterozygous and one homozygous carrier were detected. The penetrance of ASCD deafness was calculated to be 0.75. As determined by Fisher's exact test, homozygosity for the *KLF7* variant was significantly associated with congenital deafness ($p = 0.014$). The odds ratio_{AA} = 6.8 (95% CI [0.68, 67.25]), i.e., homozygous carriers are 6.8 times more likely to be deaf than wild type.

Table 5. Genotype distribution of *KLF7* variant in 31 deaf and 28 normal hearing ASCDs dogs.

Phenotype	G_G	A_G	A_A	Total Number	<i>p</i> ^(c)
Unilaterally deaf	10	10	1	21	0.054
Bilaterally deaf	3	4	3	10	0.010
Deafness (uni ^(a) + bi ^(b))	13	14	4	31	0.014
Normal hearing	22	5	1	28	

^(a) uni: Unilaterally deaf; ^(b) bi: Bilaterally deaf; ^(c) *p*-value using Fisher's exact test.

4. Discussion

Deafness is a common disorder among dogs, and the observed prevalence is highest in Dalmatians (29.9%) [3] and 17.8% in ASCD [12]. Even selective breeding based on deafness phenotyping decreased the prevalence in Dalmatians only to 17.8% [40]. Several other dog breeds also show rather high prevalence rates (>10%), e.g., Australian Cattle Dog and Bull Terrier [3]. To accelerate the decline of overall prevalence of congenital sensorineural deafness, it would be important to identify the genetic cause of the disorder to enable informed breeding.

We used four ASCD DNA samples from a previous study of deafness in Australian Stumpy Tail Cattle Dogs for GWAS and WGS analysis. The previous study used a genome screen with 325 microsatellite (290 were used for linkage mapping) to determine a significantly linked deafness region on CFA10 [12]. However, *SOX10*, the only potential candidate gene in this region, had to be excluded, as it did not harbor any causative variants. Another promising candidate in the CFA10 region, i.e., Trio- and f-actin-binding protein (*TRIOBP*), had also to be excluded. In the above mentioned study, deafness was reported to be autosomal recessive inherited with incomplete penetrance [12]. As shown before, GWAS with multiple breeds can improve the accuracy of causative variant mapping [41,42]. Our analysis provided evidence for at least six highly associated chromosomal regions. However, due to the small number of affected dogs, some associated regions might have resulted from the close relationship of the dogs. This can be seen in the QQ-plot which showed convincing evidence for an association with some indication of a population substructure. In our study, more than half of the significant associated SNPs (7 out of 13) were located on CFA37, including the most significantly related SNP (chr37:44793, $p = 9.54 \times 10^{-21}$). In a recent study of Dalmatian deafness, signals were also detected in this region [17]. However, there was no associated peak on CFA37 reported in the previous microsatellite-based study in ASCD. A possible explanation could be that there were only five microsatellite markers on CFA37, one of which had a low degree of polymorphism (3 alleles, PIC 0.5) [12]. This might have been insufficient to detect associations on this chromosome. An alternative explanation is that ASCD deafness may be heterogeneous. There may be more than one variant causing congenital deafness in this breed, and using limited family associations may reveal private mutations. Further genotyping analysis in a wider range of affected (28) and unaffected (27) ASCDs revealed that the *KLF7* missense variant was still significantly associated with congenital deafness (Table 5). Furthermore, the penetrance of deafness in ASCD calculated based on the *KLF7* variant was 0.75, which was in agreement with the previously calculated penetrance of 0.72 [12]. Altered allele (A) frequency is 24.58% (Table 5). If we take penetrance into consideration, the deafness frequency is $(24.58\% \times 0.75) = 18.4\%$, which is also close to the previous investigation of 17.8% overall ASCD breed deafness frequency [12]. Several homozygous wild type individuals were detected among the deaf ASCDs suggesting additional genetic risk factors.

This was not surprising, as canine congenital deafness seems to be a complex disorder and different regions were detected in other GWASs for deafness so far [17].

According to our GWAS, functional relationships with deafness of genes near the significantly associated loci on most chromosomes were unapparent (Table 2). Only the region on CFA37 was further supported by WGS. In the initial GWAS 651 variants on chromosome 37 (between CFA37:7217 to CFA37:30803691) were identified (Figure 1). Variant CFA37:15503029T>C with a p -value of 8.61×10^{-6} was only 12,534 bp distant from *KLF7*. To evaluate LD over-pruning and potential effects on resolution, we repeated the GWAS using less stringent pruning parameters ($-indep\ 1000\ 5\ 4$). This increased the number of associated variants to 60,746. In agreement with the previous analysis, a variant with $-\log_{10}p\text{-value} = 14.68$ at position CFA37:15463045 remained in the vicinity of *KLF7* (Table S6) and a significantly associated region spanning from CFA37:15463045 to CFA37:16433709 was detected harboring *KLF7* (CFA37:15515563–15607345). As expected, a further reduction of pruning stringency resulted in more chromosomal regions above the significant threshold (Figure S1). However, especially on CFA10, no significantly associated variants were identified.

In addition, we applied whole genome sequencing of the deaf dogs and used a large number of available canine whole genome sequence data as controls to improve the accuracy and efficiency of causative variant identification. Several GWAS of canine complex hereditary deafness failed to identify causative variants with the exception of two associated genes (*MYO7A*, *PTPRQ*) causative for a specific form of canine congenital bilateral deafness with vestibular disease [14,15].

For next generation sequence analysis in the present study, functional variants within coding regions were primarily considered due to their direct impact on protein function [43]. We filtered all variants using an autosomal recessive model, however, no functional variants fulfilled this mode of inheritance. Again, the chromosomal region 1 Mb up- and downstream of *SOX10* (CFA10:25680441–27690530) was checked using IGV, but no deafness associated variants including larger structural variants were identified. After WGS analysis and variant effect prediction, only two missense variants within *HEG1* and *KLF7* remained. *HEG1* is involved in cardiovascular development [44] and therefore seemed unlikely to be involved in the development of deafness. However, the candidate variant (NC_006619.3: g.15562684G>A) in *KLF7* (CFA37:15515563–15607345) was close to the significantly associated SNP CFA37:16399127 ($p = 2.66 \times 10^{-8}$) (Table 2). *KLF7* is a zinc finger transcription factor and has been reported to play a role in the nervous system and is vital for neuronal morphogenesis that could function in axon outgrowth [18]. *KLF7* was suggested to have potential functions in neurogenesis of mice, like neuronal differentiation and maturation [45]. *KLF7* was also found to promote axon regeneration [46]. Furthermore, *KLF7* is required for the development of sensory neurons [47], and it has been reported to play roles in neurotransmission and synaptic vesicle trafficking [48]. These two processes have important influences on the auditory system, and therefore disruption of *KLF7* could lead to hearing impairment and dysfunction [49]. Indeed, *KLF7* was confirmed to be expressed in the otic placode which will develop into ears, indicating *KLF7* could have an effect on ear development [19]. *KLF7* was also detected to be a fibroblast growth factor (FGF) responsive factor in ear progenitor induction processes, which implies it may be involved in early ear induction [50]. *KLF7* has been considered as one high quality candidate gene for human branchio-oto-renal syndrome, which is an autosomal dominant disease with hearing loss as one clinical sign [51]. *KLF7* was the nearby gene (50,519 bp distance) of one significant signal in adult hearing difficulty GWAS [52]. One recent GWAS of hearing-related traits with up to 330,759 individuals (UK Biobank) revealed 31 significant genomic risk loci for adult hearing difficulty, *KLF7* was also detected to be significantly associated [53]. Furthermore, the protein sequence segments surrounding *KLF7* variant are much more conserved than that of *HEG1* among the same 7 species (Figure 2). Recently, *KLF7* has been reported to directly regulate GATA Binding Protein 3 (*GATA3*) expression [54]. *GATA3* is expressed in the otic placode and is involved in inner ear development [55]. Though

the interaction between *KLF7* and *GATA3* was reported in chicken adipogenesis, *KLF7* is quite conserved among several species (Figure 2). Knockdown of Paired Box Protein Pax-2 (*PAX2*) (inner ear development gene) led to a significant up-regulation of both *KLF7* and *GATA3* expression [19], which implies *KLF7* and *GATA3* are probably involved in the same pathway. Furthermore, *GATA3* is the causative gene for human hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome [56]. Therefore, *KLF7* could interact with *GATA3* during the development of inner ear, and defects in *KLF7* could affect *GATA3* normal expression patterns in otic placode. This may be a potential cause of hearing loss in ASCD cases. The incomplete penetrance presented by the *KLF7* variant in deafness may be related to its role as a transcription factor that is involved in a specific part of the hearing pathway. Our findings could provide clues for the functional analysis of the *KLF7* in inner ear development. Functional analysis of *KLF7* regarding ear development may provide further evidence for its role in deafness.

Another intriguing possible pathway is suggested by the finding of a KLF binding site upstream of the *M* promoter of Microphthalmia-associated transcription factor (*M-MITF*) that induces gene expression changes in humans [57]. Although the aforementioned study was related to melanoma development, *M-MITF* has been identified as the locus responsible for white coat patterning in dogs [58]. Hereditary deafness has been reported to be associated with white pigmentation in several species, e.g., by affecting *M-MITF* isoform expression in pigs [59] and cows [60] as well as humans [61]. Canine deafness was also linked with white pigmentation due to the merle and piebald locus [62]. Congenital sensorineural deafness of English Bull Terrier is predominant in individuals with white coat color [63]. Similarly, congenital hereditary sensorineural deafness in the Australian Cattle Dog was negatively associated with bilateral facial masks, also individuals with pigmented body patches showed a lower risk of deafness [64]. An inverse association of pigmented head patches and congenital sensorineural deafness was also observed in Dalmatians, while on the other hand, a positive correlation was detected with blue irises [65–71]. In ASCD, congenital sensorineural deafness was moderately significant associated with red/blue coat color, but not with speckling and facial masks [12]. However, no functional alterations in genes related to coat color or pigmentation were detected after filtering for case–control setting in the present study. Thus far, no causative variants within genes involved in pigmentation have been identified in canine deafness. Some pigmentation genes have actually been excluded as candidates in different dog breeds, e.g., c-Kit (*KIT*) and melanocyte protein 17 (*SILV*) [72,73]. An alternative explanation is that deafness caused by dysfunctions of other biological processes may be more common, such as ear development and morphogenesis. This is highly relevant in the Gene Ontology (GO) category analysis of potential canine hereditary deafness genes [2]. In our study, *KLF7* was reported to participate in inner ear development processes [50]. There is good evidence here that the *KLF7* variant contributes to deafness, but the genotyping data supports the view that this is a multigenic/multifactorial disease, and so this is one contributing mutation.

5. Conclusions

In summary, a missense variant within *KLF7* gene has been identified to be significantly associated with congenital deafness in Australian Stumpy Tail Cattle Dogs. As *KLF7* gene was reported to be expressed in the inner ear and associated with human hearing difficulties, our findings could provide clues for further elucidating novel genetic causes for human hearing loss.

Supplementary Materials: The following are available online at <https://www.mdpi.com/2073-4425/12/4/467/s1>, Figure S1. Manhattan and QQ plots of the Genome Wide Association Analysis (GWAS) for ASCD deafness. Table S1: *KLF7* variant genotypes of 59 ASCDs. Table S2: Dog breeds of herding group. Table S3: Human deafness related genes. Table S4: Variants of recessive inheritance model and annotated with Ensembl release 101. Table S5: Private variants and annotated with Ensembl release 101. Table S6: Significantly associated SNPs above Bonferroni significance threshold (6.78).

Author Contributions: F.X.: Investigation, visualization, writing—original draft. S.S. (Shuwen Shan): Investigation, visualization, writing—original draft. S.S. (Susan Sommerlad): Resources, writing—review and editing. J.M.S.: Resources, writing—review and editing. B.B.: Conceptualization, resources, project administration, funding acquisition, supervision, writing—original draft, review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The collection of dog blood samples was done by S. Sommerlad at the time of BAER testing. The collection of samples was approved by the “Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit” (33.19-42502-05-15A506) according to §8a Abs. 1 Nr. 2 of the TierSchG. All ASCDs were tested and sampled under approval of The University of Queensland’s Animal Ethics Committee.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are openly available in Center for Open Science (OSF) reference number <https://osf.io/z36ap> (accessed on 24 March 2021).

Acknowledgments: S. Shan and F. Xu were supported by a fellowship of the China Scholarship Council (CSC). The authors would like to thank Dogs Queensland and the participating Australian Stumpy Tail Cattle Dog breeders of Queensland and New South Wales Australia, for their cooperation.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rak, S.G.; Distl, O. Congenital sensorineural deafness in dogs: A molecular genetic approach toward unravelling the responsible genes. *Vet. J.* **2005**, *169*, 188–196. [[CrossRef](#)] [[PubMed](#)]
2. Kelly-Smith, M.; Strain, G.M. STRING Data Mining of GWAS Data in Canine Hereditary Pigment-Associated Deafness. *Vet. Anim. Sci.* **2020**, *9*, 100118. [[CrossRef](#)] [[PubMed](#)]
3. Strain, G.M. Deafness prevalence and pigmentation and gender associations in dog breeds at risk. *Vet. J.* **2004**, *167*, 23–32. [[CrossRef](#)]
4. Strain, G.M. The genetics of deafness in domestic animals. *Front. Vet. Sci.* **2015**, *2*, 29. [[CrossRef](#)]
5. Strain, G.M. Aetiology, prevalence and diagnosis of deafness in dogs and cats. *Br. Vet. J.* **1996**, *152*, 17–36. [[CrossRef](#)]
6. Cunningham, L.L.; Tucci, D.L. Hearing loss in adults. *N. Engl. J. Med.* **2017**, *377*, 2465–2473. [[CrossRef](#)]
7. Hiraide, F.; Paparella, M.M. Histopathology of the temporal bones of deaf dogs. *Auris Nasus Larynx* **1988**, *15*, 97–104. [[CrossRef](#)]
8. Coppens, A.; Resibois, A.; Poncelet, L. Bilateral deafness in a maltese terrier and a great pyrenean puppy: Inner ear morphology. *J. Comp. Pathol.* **2000**, *122*, 223–228. [[CrossRef](#)] [[PubMed](#)]
9. Strain, G.M. Hereditary deafness in dogs and cats: Causes, prevalence, and current research. *Strain* **2003**, *225*, 578–9758.
10. Sims, M.H. Electrodiagnostic evaluation of auditory function. *Vet. Clin. N. Am. Small Anim. Pract.* **1988**, *18*, 913–944. [[CrossRef](#)]
11. Wilson, W.J.; Mills, P.C. Brainstem auditory-evoked response in dogs. *Am. J. Vet. Res.* **2005**, *66*, 2177–2187. [[CrossRef](#)]
12. Sommerlad, S.; McRae, A.F.; McDonald, B.; Johnstone, I.; Cuttall, L.; Seddon, J.M.; O’Leary, C.A. Congenital sensorineural deafness in Australian stumpy-tail cattle dogs is an autosomal recessive trait that maps to CFA10. *PLoS ONE* **2010**, *5*, e13364. [[CrossRef](#)] [[PubMed](#)]
13. Yokoyama, J.S.; Lam, E.T.; Ruhe, A.L.; Erdman, C.A.; Robertson, K.R.; Webb, A.A.; Williams, D.C.; Chang, M.L.; Hytönen, M.K.; Lohi, H. Variation in genes related to cochlear biology is strongly associated with adult-onset deafness in border collies. *PLoS Genet.* **2012**, *8*, e13364. [[CrossRef](#)]
14. Guevar, J.; Olby, N.J.; Meurs, K.M.; Yost, O.; Friedenberg, S.G. Deafness and vestibular dysfunction in a Doberman Pinscher puppy associated with a mutation in the PTPRQ gene. *J. Vet. Intern. Med.* **2018**, *32*, 665–669. [[CrossRef](#)]
15. Webb, A.A.; Ruhe, A.L.; Neff, M.W. A missense mutation in MYO7A is associated with bilateral deafness and vestibular dysfunction in the Doberman pinscher breed. *Can. J. Vet. Res.* **2019**, *83*, 142–148.
16. Kluth, S.; Distl, O. Congenital sensorineural deafness in Dalmatian dogs associated with quantitative trait loci. *PLoS ONE* **2013**, *8*, e1002898. [[CrossRef](#)]
17. Hayward, J.J.; Kelly-Smith, M.; Boyko, A.R.; Burmeister, L.; De Risio, L.; Mellersh, C.; Freeman, J.; Strain, G.M. A genome-wide association study of deafness in three canine breeds. *PLoS ONE* **2020**, *15*, e0232900. [[CrossRef](#)]
18. Laub, F.; Lei, L.; Sumiyoshi, H.; Kajimura, D.; Dragomir, C.; Smaldone, S.; Puche, A.C.; Petros, T.J.; Mason, C.; Parada, L.F. Transcription factor KLF7 is important for neuronal morphogenesis in selected regions of the nervous system. *Mol. Cell. Biol.* **2005**, *25*, 5699–5711. [[CrossRef](#)]
19. Chen, J.; Tambalo, M.; Barembaum, M.; Ranganathan, R.; Simões-Costa, M.; Bronner, M.E.; Streit, A. A systems-level approach reveals new gene regulatory modules in the developing ear. *Development* **2017**, *144*, 1531–1543. [[CrossRef](#)]
20. Strain, G. Brainstem auditory evoked response (BAER). *Deaf. Dogs Cats* **2011**, 83–107. [[CrossRef](#)]

21. Plassais, J.; Kim, J.; Davis, B.W.; Karyadi, D.M.; Hogan, A.N.; Harris, A.C.; Decker, B.; Parker, H.G.; Ostrander, E.A. Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* **2019**, *10*, 1–14. [\[CrossRef\]](#)
22. Jagannathan, V.; Drögemüller, C.; Leeb, T.; Consortium, D.B.V.D.; Aguirre, G.; André, C.; Bannasch, D.; Becker, D.; Davis, B.; Ekenstedt, K. A comprehensive biomedical variant catalogue based on whole genome sequences of 582 dogs and eight wolves. *Anim. Genet.* **2019**, *50*, 695–704. [\[CrossRef\]](#)
23. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [\[CrossRef\]](#)
24. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [\[CrossRef\]](#)
25. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [\[CrossRef\]](#)
26. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [\[CrossRef\]](#)
27. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; De Bakker, P.I.; Daly, M.J. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [\[CrossRef\]](#)
28. Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **2006**, *38*, 904–909. [\[CrossRef\]](#)
29. Zhou, X.; Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **2012**, *44*, 821–824. [\[CrossRef\]](#)
30. Turner, S.D. qqman: An R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv* **2014**, 005165. [\[CrossRef\]](#)
31. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Thorvaldsdóttir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **2013**, *14*, 178–192. [\[CrossRef\]](#)
33. Kumar, P.; Henikoff, S.; Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **2009**, *4*, 1073. [\[CrossRef\]](#)
34. Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **2013**, *76*, 7–20. [\[CrossRef\]](#)
35. Choi, Y.; Chan, A.P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **2015**, *31*, 2745–2747. [\[CrossRef\]](#)
36. Clarke, G.M.; Anderson, C.A.; Pettersson, F.H.; Cardon, L.R.; Morris, A.P.; Zondervan, K.T. Basic statistical analysis in genetic case-control studies. *Nat. Protoc.* **2011**, *6*, 121–133. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Jourquin, J.; Duncan, D.; Shi, Z.; Zhang, B. GLAD4U: Deriving and prioritizing gene lists from PubMed literature. *BMC Genom.* **2012**, *13*, S20. [\[CrossRef\]](#)
38. Taylor, W. Residual colours: A proposal for aminochromography. *Protein Eng.* **1997**, *10*, 743–746. [\[CrossRef\]](#)
39. Schneider, T.D.; Stephens, R.M. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **1990**, *18*, 6097–6100. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Lewis, T.; Freeman, J.; De Risio, L. Decline in prevalence of congenital sensorineural deafness in Dalmatian dogs in the United Kingdom. *J. Vet. Intern. Med.* **2020**, *34*, 1524–1531. [\[CrossRef\]](#)
41. Daetwyler, H.D.; Capitan, A.; Pausch, H.; Stothard, P.; Van Binsbergen, R.; Brøndum, R.F.; Liao, X.; Djari, A.; Rodriguez, S.C.; Grohs, C. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **2014**, *46*, 858. [\[CrossRef\]](#)
42. Raven, L.-A.; Cocks, B.G.; Hayes, B.J. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genom.* **2014**, *15*, 62. [\[CrossRef\]](#)
43. Yan, J.; Wang, X. Detection of Disease-associated Mutations and Biomarkers Using Next-generation Sequencing. *Detect. Methods Precis. Med.* **2020**, *18*, 119.
44. Donat, S.; Lourenço, M.; Paolini, A.; Otten, C.; Renz, M.; Abdelilah-Seyfried, S. Heg1 and Ccm1/2 proteins control endocardial mechanosensitivity during zebrafish valvulogenesis. *Elife* **2018**, *7*, e28939. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Laub, F.; Aldabe, R.; Friedrich Jr, V.; Ohnishi, S.; Yoshida, T.; Ramirez, F. Developmental expression of mouse Krüppel-like transcription factor KLF7 suggests a potential role in neurogenesis. *Dev. Biol.* **2001**, *233*, 305–318. [\[CrossRef\]](#)
46. Blackmore, M.G.; Wang, Z.; Lerch, J.K.; Motti, D.; Zhang, Y.P.; Shields, C.B.; Lee, J.K.; Goldberg, J.L.; Lemmon, V.P.; Bixby, J.L. Krüppel-like Factor 7 engineered for transcriptional activation promotes axon regeneration in the adult corticospinal tract. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 7517–7522. [\[CrossRef\]](#)
47. Lei, L.; Laub, F.; Lush, M.; Romero, M.; Zhou, J.; Luikart, B.; Klesse, L.; Ramirez, F.; Parada, L.F. The zinc finger transcription factor Klf7 is required for TrkA gene expression and development of nociceptive sensory neurons. *Genes Dev.* **2005**, *19*, 1354–1364. [\[CrossRef\]](#)

48. Kajimura, D.; Dragomir, C.; Ramirez, F.; Laub, F. Identification of genes regulated by transcription factor KLF7 in differentiating olfactory sensory neurons. *Gene* **2007**, *388*, 34–42. [[CrossRef](#)] [[PubMed](#)]
49. Akçimen, F.; Vural, A.; Durmuş, H.; Çakar, A.; Houlden, H.; Parman, Y.G.; Başak, A.N. A novel homozygous FBXO38 variant causes an early-onset distal hereditary motor neuropathy type IID. *J. Hum. Genet.* **2019**, *64*, 1141–1144. [[CrossRef](#)] [[PubMed](#)]
50. Tambalo, M.; Anwar, M.; Ahmed, M.; Streit, A. Enhancer activation by FGF signalling during otic induction. *Dev. Biol.* **2020**, *457*, 69–82. [[CrossRef](#)]
51. Brophy, P.D.; Alasti, F.; Darbro, B.W.; Clarke, J.; Nishimura, C.; Cobb, B.; Smith, R.J.; Manak, J.R. Genome-wide copy number variation analysis of a Branchio-oto-renal syndrome cohort identifies a recombination hotspot and implicates new candidate genes. *Hum. Genet.* **2013**, *132*, 1339–1350. [[CrossRef](#)]
52. Wells, H.R.; Freidin, M.B.; Abidin, F.N.Z.; Payton, A.; Dawes, P.; Munro, K.J.; Morton, C.C.; Moore, D.R.; Dawson, S.J.; Williams, F.M. GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. *Am. J. Hum. Genet.* **2019**, *105*, 788–802. [[CrossRef](#)]
53. Kalra, G.; Milon, B.; Casella, A.M.; Herb, B.R.; Humphries, E.; Song, Y.; Rose, K.P.; Hertzano, R.; Ament, S.A. Biological insights from multi-omic analysis of 31 genomic risk loci for adult hearing difficulty. *PLoS Genet.* **2020**, *16*, e1009025. [[CrossRef](#)]
54. Sun, Y.; Jin, Z.; Zhang, X.; Cui, T.; Zhang, W.; Shao, S.; Li, H.; Wang, N. GATA binding protein 3 is a direct target of Kruppel-like transcription factor 7 and inhibits chicken adipogenesis. *Front. Physiol.* **2020**, *11*, 610. [[CrossRef](#)]
55. Lawoko-Kerali, G.; Rivolta, M.N.; Holley, M. Expression of the transcription factors GATA3 and Pax2 during development of the mammalian inner ear. *J. Comp. Neurol.* **2002**, *442*, 378–391. [[CrossRef](#)] [[PubMed](#)]
56. Van Esch, H.; Groenen, P.; Nesbit, M.A.; Schuffenhauer, S.; Lichtner, P.; Vanderlinden, G.; Harding, B.; Beetz, R.; Bilous, R.W.; Holdaway, I. GATA3 haplo-insufficiency causes human HDR syndrome. *Nature* **2000**, *406*, 419–422. [[CrossRef](#)] [[PubMed](#)]
57. Pierrat, M.-J.; Marsaud, V.; Mauviel, A.; Javelaud, D. Expression of microphthalmia-associated transcription factor (MITF), which is critical for melanoma progression, is inhibited by both transcription factor GLI2 and transforming growth factor- β . *J. Biol. Chem.* **2012**, *287*, 17996–18004. [[CrossRef](#)]
58. Karlsson, E.K.; Baranowska, I.; Wade, C.M.; Hillbertz, N.H.S.; Zody, M.C.; Anderson, N.; Biagi, T.M.; Patterson, N.; Pielberg, G.R.; Kulbokas, E.J. Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.* **2007**, *39*, 1321–1328. [[CrossRef](#)] [[PubMed](#)]
59. Chen, L.; Guo, W.; Ren, L.; Yang, M.; Zhao, Y.; Guo, Z.; Yi, H.; Li, M.; Hu, Y.; Long, X. A de novo silencer causes elimination of MITF-M expression and profound hearing loss in pigs. *BMC Biol.* **2016**, *14*, 1–15. [[CrossRef](#)] [[PubMed](#)]
60. Philipp, U.; Lupp, B.; Mömke, S.; Stein, V.; Tipold, A.; Eule, J.C.; Rehage, J.; Distl, O. A MITF mutation associated with a dominant white phenotype and bilateral deafness in German Fleckvieh cattle. *PLoS ONE* **2011**, *6*, e28857. [[CrossRef](#)]
61. Tassabehji, M.; Newton, V.E.; Read, A.P. Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nat. Genet.* **1994**, *8*, 251–255. [[CrossRef](#)]
62. Strain, G.M. Canine deafness. *Vet. Clin. Small Anim. Pract.* **2012**, *42*, 1209–1224. [[CrossRef](#)]
63. De Risio, L.; Freeman, J.; Lewis, T. Prevalence, heritability and genetic correlations of congenital sensorineural deafness and coat pigmentation phenotype in the English bull terrier. *BMC Vet. Res.* **2016**, *12*, 146. [[CrossRef](#)] [[PubMed](#)]
64. Sommerlad, S.F.; Morton, J.M.; Haile-Mariam, M.; Johnstone, I.; Seddon, J.M.; O’Leary, C.A. Prevalence of congenital hereditary sensorineural deafness in Australian Cattle Dogs and associations with coat characteristics and sex. *BMC Vet. Res.* **2012**, *8*, 202. [[CrossRef](#)]
65. Strain, G.M.; Kearney, M.T.; Gignac, I.J.; Levesque, D.C.; Nelson, H.J.; Tedford, B.L.; Remsen, L.G. Brainstem auditory-evoked potential assessment of congenital deafness in Dalmatians: Associations with phenotypic markers. *J. Vet. Intern. Med.* **1992**, *6*, 175–182. [[CrossRef](#)]
66. Greibrokk, T. Hereditary deafness in the Dalmation: Relationship to eye and coat color. *Journal* **1995**, *30*, 170–176.
67. Famula, T.; Oberbauer, A.; Sousa, C. A threshold model analysis of deafness in Dalmatians. *Mamm. Genome* **1996**, *7*, 650–653. [[CrossRef](#)] [[PubMed](#)]
68. Wood, J.; Lakhani, K. Prevalence and prevention of deafness in the Dalmatian—Assessing the effect of parental hearing status and gender using ordinary logistic and generalized random litter effect models. *Vet. J.* **1997**, *154*, 121–133. [[CrossRef](#)]
69. Muhle, A.C.; Jaggy, A.; Stricker, C.; Steffen, F.; Dolf, G.; Busato, A.; Kornberg, M.; Mariscoli, M.; Srenk, P.; Gaillard, C. Further contributions to the genetic aspect of congenital sensorineural deafness in Dalmatians. *Vet. J.* **2002**, *163*, 311–318. [[CrossRef](#)] [[PubMed](#)]
70. Juraschko, K.; Meyer-Lindenberg, A.; Nolte, I.; Distl, O. Analysis of systematic effects on congenital sensorineural deafness in German Dalmatian dogs. *Vet. J.* **2003**, *166*, 164–169. [[CrossRef](#)]
71. Cargill, E.; Famula, T.; Strain, G.; Murphy, K. Heritability and segregation analysis of deafness in US Dalmatians. *Genetics* **2004**, *166*, 1385–1393. [[CrossRef](#)] [[PubMed](#)]
72. Metallinos, D.; Rine, J. Exclusion of EDNRB and KIT as the basis for white spotting in Border Collies. *Genome Biol.* **2000**, *1*, 1–4. [[CrossRef](#)] [[PubMed](#)]
73. Stritzel, S.; Wohlke, A.; Distl, O. Elimination of SILV as a candidate for congenital sensorineural deafness in Dalmatian dogs. *Anim. Genet.* **2007**, *38*, 662. [[CrossRef](#)] [[PubMed](#)]

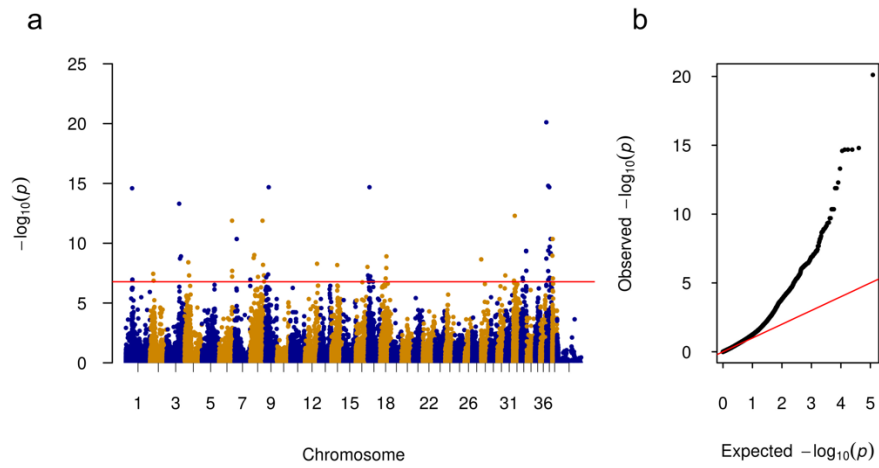


Figure S1. Manhattan and QQ plots of the Genome Wide Association Analysis (GWAS) for ASCD deafness. **(a)** Association of 60,746 variants with bilateral deafness in 47 dogs (3 cases and 44 controls). The plot shows the $-\log_{10}p$ -values for all variants. 60,746 were remained after pruning by Linkage Disequilibrium with parameters `--indep 1000 5 4`. The red horizontal line represents the Bonferroni genome-wide significance threshold of $-\log_{10}(0.01/60,746) = 6.78$. **(b)** Quantile-quantile (QQ) plot showed the observed $-\log_{10}p$ -values in the black curve, the red line indicated the distribution of expected $-\log_{10}p$ -values. λ is calculated to be 1.10.

CHAPTER 5

General Discussion

1 Significance of the research

This thesis describes potential candidate variants and genes for three disorders, namely an idiopathic infertility in Holstein bull, hemophilia B in Hovawart and congenital deafness in ASCD, and attempts to shed light on their causal principles from a genetic perspective.

Focusing on bull fertility is an important issue in cattle breeding, especially in the context of rapidly declining fertility due to intense selection for productive traits. Cattle are an important component of farm animal, primarily providing milk for humans. Therefore, identifying a candidate causal variant in *ABHD16B* for idiopathic infertility in bull can improve cattle breeding by controlling the proportion of unfavorable variants in the herd.

Hemophilia B is a recessive, X-linked bleeding disorder that severely affects the normal life of patients and the welfare of dogs. We have elucidated the causative variant of a hemophilia B Leyden found in a Hovawart family, providing theoretical support for overcoming canine hemophilia B.

We have identified a potential candidate variant in *KLF7* that has a penetrance of 0.75 for the congenital deafness in ASCD. Dogs are good models for complex disease studies, such as deafness with high heterogeneity, because dogs have a simplified genetic background and high linkage disequilibrium (LD).

2 Biological studies in animals help humans understand themselves

Within the past 20 years, the field of genetics and genomics has undergone extraordinary changes due to the development of new technologies, large-scale resources, and computational approaches [1]. The year 2001 marked a new milestone with the publication of the sketch of the human genome [2]. By 2003, the human genome was declared complete, yet there were still many gaps that had not been filled [3]. Initial analysis revealed that about 50% of the human genome was made up of repetitive sequences [4]. Scientists analyzed the coding protein sequences by RNA expression data and found about 40,000 protein-coding genes [5]. Subsequently, the genomes of mammals such as mouse [6], rat [7] and dog [8] were published and scientists realized that gene function could be better analyzed using comparative genomics [1]. For example, using conserved synthesis among four mammalian species - human, mouse, rat and dog - scientists revised the number of mammalian genes to approximately 20,000 [8]. Similarly, disease-related genes identified in cattle and dogs could

inform the resolution of similar diseases in humans. There are more than 350 canine diseases that are analogous to humans [9]. Some of these diseases are caused by the same abnormal gene products, such as enzyme, receptor, transporter or structural protein [10]. Many diseases similar to humans have been identified in dogs, such as canine systemic lupus erythematosus [11], skeletal developmental disorders [12], osteosarcoma [13], and canine compulsive disorder [14, 15].

3 Candidate Gene Targeting

3.1 Phenotype definition and measurement

Detailed phenotype description, rigorous phenotype definition and classification are the basis of biological research, especially for the complex traits. Some phenotypic heterogeneity (diverse disease manifestations) is caused by genetic heterogeneity (different sets of genetic factors for each disease subtype). Careful selection of cases can minimize the impact of misclassification due to possible phenotypic heterogeneity, and enrich the genetic components of sub-phenotype. This step can improve the ability to detect genetic variants associated with specific sub-phenotypes [16].

Male infertility is a multifactorial pathological condition with a highly complex genetic background. The histological phenotypes of semen and testis exhibit great heterogeneity, and about 2,000 genes are involved in human spermatogenesis [17]. Common factors that contribute to male infertility include azoospermia, oligospermia, asthenospermia, and abnormal morphology of testis and spermatozoa, which account for a substantial fraction of causative factors (reviewed by [17]). However, there are still many cases with unknown etiology. Therefore, discovering novel genetic factors in idiopathic infertility is a key step to improve male fertility. In dairy cattle breeding, much less attention has been paid to improving bull fertility than to improving cow fertility. But the fact is, widespread dissemination of semen among different populations may amplify the impact of individual bulls on profitability. In some cases, semen companies may adjust the number of sperms in the artificial insemination device to compensate for reduced fertility or suboptimal semen parameters [18]. This results in a wide range of fertility in the bull used for breeding. Therefore, semen assessment needs to be improved [19].

3.2 GWAS and WGS

The first genome-wide association study was published in 2005, which identified the association of complement factor H with age-related macular degeneration [20]. This study prompted the genetics community to recognize the efficiency and feasibility of GWAS in finding variants associated with unknown diseases. With the following explosion of GWAS publications, the genetics studies of human complex diseases have entered a new era, and then expanded to the field of animals and plants. With the release of the first *Bos taurus* whole genome assembly in 2009 [21] and the availability of commercial genotyping assay [22], a large number of GWAS have emerged in the bovine setting. The importance of bull's contribution to reproductive success has been fully demonstrated [23, 24]. However, compared with female fertility research, there are very few studies on quantitative trait loci (QTL), GWAS or candidate genes for male fertility. Some reviews have summed up data that associated with bull fertility traits [25, 26]. Fonseca et al. [26] systematically reviewed the genetic mechanisms of bull fertility traits. Spermatogenic and testicular related traits show a high genetic heterogeneity. GWAS related to testicular traits are mostly carried out in beef, and more GWAS studies in dairy cattle are about sperm traits. Compared with spermatogenic traits, testicular traits have higher heritability values, easier measurement methods and higher correlation with fertility traits, indicating that testicular measurements are a more stable and representative measure of beef and dairy cattle fertility. Thus, more GWASs using testicular features should be implemented in dairy cattle. In addition to the dominant sperm infertility issues, such as motility and morphology defects, idiopathic male subfertility has also been reported in Fleckvieh, caused by a loss-of-function mutation in *TMEM95* [27]. In chapter 2, we described a case of idiopathic male infertility in Holstein cattle with normal sperm characteristics but infertility. GWAS analysis determined that the idiopathic infertility was related to a region on BTA13. Therefore, AI bulls provide an ideal model system to study the relationship between male fertility and genome-wide variation because of their complete pedigree records, extensive breeding records and offspring records [28].

The application of GWAS also has limitations. Compared with rare SNPs, GWAS is easier to detect common SNPs with higher statistical power. In addition, because the selection of SNPs is guided by LD information rather than function, it is more likely that the SNPs identified by GWAS are only surrogate markers for functional variation [16]. Besides, the applicability of GWAS results in different populations depends on characters of markers, like allelic frequencies and LD pattern. Integrating evidence from multiple sources makes it easier to find

areas related to interesting traits that are unique to a specific group or shared among multiple groups [26]. In the future, GWAS studies can increase the sample sizes, conduct larger studies or combine the existing GWAS data set in meta-analyses [16].

Although GWAS is a powerful tool for identifying genomic regions related to phenotypes, it lacks the ability to detect causal variations [29]. As technology advances, WGS becomes cheaper and more readily available. Therefore, it can be used to further identify candidate disease-causal variants and genes. The depth of coverage and the number of subjects to be sequenced need to be considered to increase the possibility of finding potential causal variants [29]. In this thesis, the identification of potential candidate variants was done with the help of WGS.

4 Post-GWAS era

4.1 Annotation of candidate variants and genes

The annotation of candidate causal variants or genes is an important step in revealing how they impact function, which requires comprehensive and highly accurate functional genomic databases. Currently, a large number of phenotype-associated regions and markers have been discovered, but their annotation and functional studies lag behind, especially for animals. Therefore, enhancing genomic annotation and increasing the available resources is an ongoing research hotspot [30]. For example, the FAANG project aims to establish the exhaustive catalogues of regulatory elements in livestock, e.g. the species-specific and tissue-specific epigenome maps [31]. FAANG's preliminary work focused on the generation of the basic data for annotation, mainly based on the individual level [32]. The Dog 10k project plans to use WGS data from 10,000 canines to improve existing reference genomes from Boxers, to create new reference genomes using other canines, and to apply these data to other scientific questions [33].

Commonly used functional analysis approaches for associated regions including identification of evolutionary conserved regions, distal regulatory elements (enhancers, silencers, etc.), promoter regions, enriched Gene Ontology (GO) terms, metabolic pathways and previously annotated related QTLs [26]. Only few risk-associated loci harbor variants that directly affect coding sequence [34]. In fact, more genetic risk variants fall within the regulatory regions of genes [35]. Some genetic risk variants can control gene expression by regulating cis-regulatory mechanisms including DNA methylation, transcription factor binding, chromatin

looping, or miRNA recruitment (reviewed by [36]). Many techniques and methods have been developed to detect regulatory elements. At first, Chromatin immunoprecipitation (ChIP) was used to detect the location of individual transcriptional regulatory sequences [37]. Over time, genome-wide location of DNA-bound proteins were identified by ChIP-on-chip and ChIP-seq [38], and they have been extended to look at differential methylation and acetylation of genomic bases and their binding proteins [39-41]. Transcriptional regulatory elements also can be recognized by assays such as histone modification or DNase sensitivity (or hypersensitivity) [39, 42-44]. Those methods provide more valuable information than studying sequence conservation, because regulatory sequences like enhancers usually lack sequence conservation [45], and show cell type [43] and species specificity [46, 47] in controlling gene expression. Meanwhile, Chromosome Conformation Capture (3C) [48] and derivative methods (4C, 5C, Hi-C and ChIA-PET) [49-54] have made it possible to detect chromosome structure in the three-dimensional space of the nucleus. Many combinations of regulatory enhancer-enhancer, enhancer-promoter or promoter-promoter have been uncovered [55-57]. Hi-C also can be used to detect the topologically associated domains of chromosome structure, which helps scientists understand the genomic composition of specific regulatory regions in specific tissues [58]. Subsequently, Assay for Transposase-accessible chromatin using sequencing (ATAC-seq) was reported to discover open chromatin [59] and was widely used for cis-regulatory sequences mapping. It was found that studies based on mixed cell populations may lead to the loss of important information, especially in tumors where genetic heterogeneity is common [60-62]. The idea of single-cell analysis is proposed, which may fundamentally change our view of how multicellular organisms work and give rise to new research questions [63]. Therefore, single-cell sequencing technologies were developed to try to understand the function of individual cells in their microenvironment. Similarly, single-cell analysis has been combined with other techniques in the study of transcriptional regulatory elements. For example, single-cell Hi-C was introduced in order to shift the probabilistic chromosome conformation averaged from millions of cells towards the resolution of chromosome and genome structure in a single cell [64]. Single-cell ATAC-seq (scATAC-seq) has been developed by combining with cell indexing or isolation methods, such as the use of scTHS-seq [65], microfluidic devices (Fluidigm, C1) [66], nano-well array (ICELL8, Takara Bio) [67] or droplet-based platforms [68, 69]. In addition, the Genotype-Tissue Expression (GTEx) project (<https://www.gtexportal.org/home/>) aims to characterize differences in gene expression levels in different individuals and different tissues and attempts to link them to genomic variation [70, 71].

4.2 Functional verification of candidate variants and genes

In addition to the high-throughput methods mentioned above, it is essential to verify the variation or gene function of a specific disease or phenotype by appropriate molecular biological methods. Functional validation of candidate variants is challenging and is often achieved with the help of model animals, as samples of cases are often difficult to obtain. However, complete experimental validation is the ultimate step to accurately define the contribution of a variant or gene to a phenotype or disease. In chapter 2, we have detected the protein expression of ABHD16B in testicular parenchyma, ductuli efferentes, and epididymis (tail, body, and head), which covers the entire process of spermatogenesis and sperm maturation. ABHD16B expression was also detected in the nucleoplasm of Leydig cells. In adults, Leydig cells increase the rate of cholesterol translocation to mitochondria, where cholesterol is metabolized to androgens [72]. Cholesterol can be derived from the *de novo* synthesis of acetate and is stored in lipid droplets [73], or from the cell membranes of Leydig cell [74]. The α/β hydrolase fold domain protein have thought to be involved in lipid metabolism and signal transduction [75]. Therefore, a hypothesis was defined that the candidate gene *ABHD16B* affects bull fertility by participating in lipid metabolism during spermatogenesis and sperm maturation. We verified this hypothesis by comparing the lipidome of the sperm plasma membrane of variant carriers and controls. The results showed that several lipid components of sperm membranes displayed significant concentration differences, suggesting the important role of *ABHD16B* in spermatogenesis and sperm maturation. During spermatogenesis, the differentiation and maturation of germ cells require extensive formation and remodeling of cell membranes. In this process, lipids undergo considerable quantitative and qualitative changes [76]. With the migration of sperm in the female reproductive tract, the sperm acquires the ability to fertilize accompanied by constant changes in lipids [77]. Several key processes such as acrosome reaction and sperm-egg fusion are membrane-related events involving exocytosis and membrane fusion [78, 79]. Lipids are multifunctional molecules in exocytosis, involved in providing platform for molecular interaction, acting as messengers in signal transduction and molecular recognition, and defining the shape of membranes [80, 81]. To our knowledge, there is no report about ABHD16B protein in reproduction. Our results showed that ABHD16B protein deficiency can lead to changes in lipid composition of the sperm plasma membrane, which may be a novel clue in fertility study.

4.3 Application and perspective

In order to meet the growing demand of the world's population for animal-based food products, scientists all over the world are working hard to increase the productivity of livestock. Breeding strategies and management practices are developed based on the accurate mapping and understanding of the causative or functional variation behind the phenotype. The rate of genetic gain in breeding is improved by increasing the accuracy of genomic selection or targeted genome editing [82].

The genomic selection is widely adopted in dairy cattle breeding and is rapidly developing in other livestock industries because it can effectively improve genetic progress [83]. Published studies have shown that the fertility traits of bulls are low to moderately heritable, indicating that the fertility of bulls can be improved through selection [23, 84-86]. For example, a recent study found that when the genomic prediction includes the markers with large effect on SCR, the predictive ability of Holstein bull fertility was improved [86]. In order to improve the prediction accuracy, genetic markers require to be carefully classified by function and more powerful and stable prediction models need to be developed [30]. Several studies have been proposed to integrate the biological information into genomic prediction, which is an effective method to improve the prediction accuracy [87-89]. Therefore, our findings can complement the bovine genome annotation information and be further used for genome-based selection. The *KLF7* variant we identified can be applied to the ASCD breeding to control the prevalence of deafness.

Genome editing technologies offer new and exciting opportunities for producing higher yielding and healthier livestock. The feasibility of genome editing in breeding is demonstrated by (1) rapid fixation of favorable alleles at segregated QTL within breeding populations [90], (2) targeted introduction of favorable alleles from other species, strains, or populations into closed breeding populations by editing [91], or (3) creation of new alleles with beneficial effects based on unbiased genome-wide screening or predictions from a priori biological knowledge of the trait in question [30, 82]. A successful example of breeding for disease resistance in domestic animals: removal of an exon of the CD163 gene in pigs gave complete resistance to reproductive and respiratory syndrome virus [92].

Furthermore, genome-editing is a good tool for fundamental research and disease treatment by editing genomes or epigenomes in cell lines, organs, and genome-scale engineering to

provide rich phenotypic and molecular models [93]. CRISPR/Cas9 system has been used for mutagenesis of specific DNA sequences to reveal the function of regulatory elements that were predicted by sequence- and chromatin-based assays [94]. CRISPR/Cas9-based acetyltransferases catalyze the acetylation of specific histones, which leads to transcriptional activation of target genes from promoters and enhancers [95]. Hemophilia, as a monogenic disease, is an ideal target for genome editing methods aimed at correcting defective genes. Gene therapy has been used in the treatment of hemophilia B [96]. Delivery of an AAV vector encoding ZFN into the mouse liver to correct the donor template to produce high levels of factor IX in a murine model of hemophilia B [97]. The development of gene editing technology may accelerate the introduction of gene therapy into the routine clinical practice.

Although technology is advancing rapidly and our understanding of biology is deepening, humanity still faces enormous challenges. We still have much work to do in food production, disease treatment, animal welfare, biodiversity conservation, and adaptation to environmental change.

Reference

1. Lindblad-Toh, K., What animals can teach us about evolution, the human genome, and human disease. *Upsala journal of medical sciences*, 2020. 125(1): p. 1-9.
2. Venter, J.C. et al., The sequence of the human genome. *science*, 2001. 291(5507): p. 1304-1351.
3. Rogers, J. *The finished genome sequence of Homo sapiens*. in *Cold Spring Harbor symposia on quantitative biology*. 2003. Cold Spring Harbor Laboratory Press.
4. Palazzo, A.F. and T.R. Gregory, The case for junk DNA. *PLoS Genet*, 2014. 10(5): p. e1004351.
5. Lander, E.S. et al., Initial sequencing and analysis of the human genome. 2001.
6. Waterston, R.H. and L. Pachter, Initial sequencing and comparative analysis of the mouse genome. *Nature*, 2002. 420(6915): p. 520-562.
7. Gibbs, R.A. and L. Pachter, Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, 2004. 428(6982): p. 493-521.
8. Lindblad-Toh, K. et al., Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 2005. 438(7069): p. 803-819.
9. Shearin, A.L. and E.A. Ostrander, Leading the way: canine models of genomics and disease. *Disease models & mechanisms*, 2010. 3(1-2): p. 27-34.
10. Ostrander, E.A., F. Galibert, and D.F. Patterson, Canine genetics comes of age. *Trends in Genetics*, 2000. 16(3): p. 117-124.
11. Wilbe, M. et al., Genome-wide association mapping identifies multiple loci for a canine SLE-related disease complex. *Nature genetics*, 2010. 42(3): p. 250.
12. Hayward, J.J. et al., Complex disease and phenotype mapping in the domestic dog. *Nature communications*, 2016. 7(1): p. 1-11.
13. Karlsson, E.K. et al., Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome biology*, 2013. 14(12): p. 1-16.
14. Dodman, N.H. et al., A canine chromosome 7 locus confers compulsive disorder susceptibility. *Molecular psychiatry*, 2010. 15(1): p. 8-10.
15. Tang, R. et al., Candidate genes and functional noncoding variants identified in a canine model of obsessive-compulsive disorder. *Genome biology*, 2014. 15(3): p. 1-15.
16. Ku, C.S. et al., The pursuit of genome-wide association studies: where are we now? *Journal of human genetics*, 2010. 55(4): p. 195-206.
17. Krausz, C. and A. Riera-Escamilla, Genetics of male infertility. *Nat Rev Urol*, 2018. 15(6): p. 369-384.
18. Saacke, R.G. et al., Relationship of seminal traits and insemination time to fertilization rate and embryo quality. *Animal Reproduction Science*, 2000. 60: p. 663-677.
19. Gross, N., F. Peñagaricano, and H. Khatib, Integration of whole-genome DNA methylation data with RNA sequencing data to identify markers for bull fertility. *Animal genetics*, 2020. 51(4): p. 502-510.
20. Klein, R.J. et al., Complement factor H polymorphism in age-related macular degeneration. *Science*, 2005. 308(5720): p. 385-389.
21. Zimin, A.V. et al., A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome biology*, 2009. 10(4): p. 1-10.
22. Matukumalli, L.K. et al., Development and characterization of a high density SNP genotyping assay for cattle. *PloS one*, 2009. 4(4): p. e5350.
23. Abdollahi-Arpanahi, R., G. Morota, and F. Penagaricano, Predicting bull fertility using genomic data and biological information. *J Dairy Sci*, 2017. 100(12): p. 9656-9666.

24. DeJarnette, J. et al., Sustaining the fertility of artificially inseminated dairy cattle: the role of the artificial insemination industry. *Journal of dairy Science*, 2004. 87: p. E93-E104.
25. Fortes, M.R. et al., Genomic regions associated with fertility traits in male and female cattle: advances from microsatellites to high-density chips and beyond. *Animal Reproduction Science*, 2013. 141(1-2): p. 1-19.
26. Fonseca, P.A.d.S. et al., Genetic mechanisms underlying spermatic and testicular traits within and among cattle breeds: Systematic review and prioritization of GWAS results. *Journal of animal science*, 2018. 96(12): p. 4978-4999.
27. Pausch, H. et al., A nonsense mutation in TMEM95 encoding a nondescript transmembrane protein causes idiopathic male subfertility in cattle. *PLoS Genet*, 2014. 10(1): p. e1004044.
28. Peddinti, D. et al., Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. *BMC systems biology*, 2008. 2(1): p. 19.
29. Freedman, M.L. et al., Principles for the post-GWAS functional characterization of cancer risk loci. *Nature genetics*, 2011. 43(6): p. 513-518.
30. Clark, E.L. et al., From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biology*, 2020. 21(1): p. 1-9.
31. Andersson, L. et al., Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome biology*, 2015. 16(1): p. 1-6.
32. Giuffra, E., C.K. Tuggle, and F. Consortium, Functional annotation of animal genomes (FAANG): current achievements and roadmap. *Annual review of animal biosciences*, 2019. 7: p. 65-88.
33. Ostrander, E.A. et al., Dog10K: an international sequencing effort to advance studies of canine domestication, phenotypes and health. *National science review*, 2019. 6(4): p. 810-824.
34. Schaub, M.A. et al., Linking disease associations with regulatory information in the human genome. *Genome research*, 2012. 22(9): p. 1748-1759.
35. Frazer, K.A. et al., Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics*, 2009. 10(4): p. 241-251.
36. Zhang, X., S.D. Bailey, and M. Lupien, Laying a solid foundation for Manhattan—‘setting the functional basis for the post-GWAS era’. *Trends in Genetics*, 2014. 30(4): p. 140-149.
37. Kim, T.H. and B. Ren, Genome-wide analysis of protein-DNA interactions. *Annu. Rev. Genomics Hum. Genet.*, 2006. 7: p. 81-102.
38. Ren, B. et al., Genome-wide location and function of DNA binding proteins. *Science*, 2000. 290(5500): p. 2306-2309.
39. Barski, A. et al., High-resolution profiling of histone methylations in the human genome. *Cell*, 2007. 129(4): p. 823-837.
40. Ramsey, S.A. et al., Genome-wide histone acetylation data improve prediction of mammalian transcription factor binding sites. *Bioinformatics*, 2010. 26(17): p. 2071-2075.
41. Visel, A. et al., ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 2009. 457(7231): p. 854-858.
42. Heintzman, N.D. et al., Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*, 2007. 39(3): p. 311-318.
43. Heintzman, N.D. et al., Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, 2009. 459(7243): p. 108-112.

44. Xi, H. et al., Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet*, 2007. 3(8): p. e136.
45. Hare, E.E. et al., *Sepsid* even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet*, 2008. 4(6): p. e1000106.
46. Villar, D. et al., Enhancer evolution across 20 mammalian species. *Cell*, 2015. 160(3): p. 554-566.
47. Arnold, C.D. et al., Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nature genetics*, 2014. 46(7): p. 685-692.
48. Dekker, J. et al., Capturing chromosome conformation. *science*, 2002. 295(5558): p. 1306-1311.
49. Zhao, Z. et al., Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra-and interchromosomal interactions. *Nature genetics*, 2006. 38(11): p. 1341-1347.
50. Simonis, M. et al., Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature genetics*, 2006. 38(11): p. 1348-1354.
51. Dostie, J. et al., Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome research*, 2006. 16(10): p. 1299-1309.
52. Lieberman-Aiden, E. et al., Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 2009. 326(5950): p. 289-293.
53. Dixon, J.R. et al., Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 2012. 485(7398): p. 376-380.
54. Fullwood, M.J. et al., An oestrogen-receptor- α -bound human chromatin interactome. *Nature*, 2009. 462(7269): p. 58-64.
55. Handoko, L. et al., CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature genetics*, 2011. 43(7): p. 630-638.
56. Sanyal, A. et al., The long-range interaction landscape of gene promoters. *Nature*, 2012. 489(7414): p. 109-113.
57. Li, G. et al., Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, 2012. 148(1-2): p. 84-98.
58. Sauerwald, N., A. Singhal, and C. Kingsford, Analysis of the structural variability of topologically associated domains as revealed by Hi-C. *NAR Genom Bioinform*, 2020. 2(1).
59. Buenrostro, J.D. et al., Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*, 2013. 10(12): p. 1213-8.
60. Park, S.Y. et al., Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. *The Journal of clinical investigation*, 2010. 120(2): p. 636-644.
61. Farabegoli, F. et al., Clone heterogeneity in diploid and aneuploid breast carcinomas as detected by FISH. *Cytometry: The Journal of the International Society for Analytical Cytology*, 2001. 46(1): p. 50-56.
62. Torres, L. et al., Intratumor genomic heterogeneity in breast cancer with clonal divergence between primary carcinomas and lymph node metastases. *Breast cancer research and treatment*, 2007. 102(2): p. 143-155.

63. Eberwine, J. et al., The promise of single-cell sequencing. *Nature Methods*, 2014. 11(1): p. 25-27.
64. Nagano, T. et al., Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 2013. 502(7469): p. 59-+.
65. Lake, B.B. et al., Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nature Biotechnology*, 2018. 36(1): p. 70-+.
66. Buenostro, J.D. et al., Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 2015. 523(7561): p. 486-U264.
67. Mezger, A. et al., High-throughput chromatin accessibility profiling at single-cell resolution. *Nature Communications*, 2018. 9.
68. Satpathy, A.T. et al., Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nature Biotechnology*, 2019. 37(8): p. 925-+.
69. Lareau, C.A. et al., Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nature Biotechnology*, 2019. 37(8): p. 916-+.
70. Aguet, F. et al., Genetic effects on gene expression across human tissues. *Nature*, 2017. 550(7675): p. 204-+.
71. Li, X. et al., The impact of rare variation on gene expression across tissues. *Nature*, 2017. 550(7675): p. 239-+.
72. Zirkin, B.R. and V. Papadopoulos, Leydig cells: formation, function, and regulation. *Biology of Reproduction*, 2018. 99(1): p. 101-111.
73. Hu, J. et al., Cellular cholesterol delivery, intracellular processing and utilization for biosynthesis of steroid hormones. *Nutrition & Metabolism*, 2010. 7.
74. Venugopal, S. et al., Plasma Membrane Origin of the Steroidogenic Pool of Cholesterol Used in Hormone-induced Acute Steroid Formation in Leydig Cells. *Journal of Biological Chemistry*, 2016. 291(50): p. 26109-26125.
75. Lord, C.C., G. Thomas, and J.M. Brown, Mammalian alpha beta hydrolase domain (ABHD) proteins: Lipid metabolizing enzymes at the interface of cell signaling and energy metabolism. *Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids*, 2013. 1831(4): p. 792-802.
76. Oresti, G.M. et al., Cell-type-specific regulation of genes involved in testicular lipid metabolism: fatty acid-binding proteins, diacylglycerol acyltransferases and perilipin 2. 2013.
77. Feki, N.C. et al., Human sperm lipid content is modified after migration into human cervical mucus. *Molecular human reproduction*, 2004. 10(2): p. 137-142.
78. Stein, K.K., P. Primakoff, and D. Myles, Sperm-egg fusion: events at the plasma membrane. *Journal of cell science*, 2004. 117(26): p. 6269-6274.
79. Stival, C. et al., Sperm capacitation and acrosome reaction in mammalian sperm. *Sperm Acrosome Biogenesis and Function During Fertilization*, 2016: p. 93-106.
80. Buffone, M.G., Sperm acrosome biogenesis and function during fertilization. 2016.
81. Tanphaichitr, N. et al., New insights into sperm-zona pellucida interaction: involvement of sperm lipid rafts. *Front Biosci*, 2007. 12: p. 1748-1766.
82. Houston, R.D. et al., Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics*, 2020. 21(7): p. 389-409.
83. Georges, M., C. Charlier, and B. Hayes, Harnessing genomic information for livestock improvement. *Nature Reviews Genetics*, 2019. 20(3): p. 135-156.
84. Utt, M.D., Prediction of bull fertility. *Animal reproduction science*, 2016. 169: p. 37-44.
85. Veerkamp, R.F. et al., Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein–Friesian cattle. *Genetics Selection Evolution*, 2016. 48(1): p. 1-14.

86. Nani, J.P., F.M. Rezende, and F. Penagaricano, Predicting male fertility in dairy cattle using markers with large effect and functional annotation data. *BMC Genomics*, 2019. 20(1): p. 258.
87. Gao, N. et al., Incorporating gene annotation into genomic prediction of complex phenotypes. *Genetics*, 2017. 207(2): p. 489-501.
88. Xiang, R. et al., Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proceedings of the National Academy of Sciences*, 2019. 116(39): p. 19398-19408.
89. Li, Z. and H. Simianer, Pan-genomic open reading frames: A potential supplement of single nucleotide polymorphisms in estimation of heritability and genomic prediction. *PLoS genetics*, 2020. 16(8): p. e1008995.
90. Jenko, J. et al., Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs (vol 47, 55, 2015). *Genetics Selection Evolution*, 2015. 47.
91. Gratacap, R.L. et al., Potential of Genome Editing to Improve Aquaculture Breeding and Production. *Trends in Genetics*, 2019. 35(9): p. 672-684.
92. Burkard, C. et al., Pigs Lacking the Scavenger Receptor Cysteine-Rich Domain 5 of CD163 Are Resistant to Porcine Reproductive and Respiratory Syndrome Virus 1 Infection (vol 92, JVI.00415-18, 2018). *Journal of Virology*, 2020. 94(15).
93. Gaj, T. et al., Genome-Editing Technologies: Principles and Applications. *Cold Spring Harbor Perspectives in Biology*, 2016. 8(12).
94. Lau, C.H. and Y. Suh, CRISPR-based strategies for studying regulatory elements and chromatin structure in mammalian gene control. *Mammalian Genome*, 2018. 29(3-4): p. 205-228.
95. Hilton, I.B. et al., Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology*, 2015. 33(5): p. 510-U225.
96. Dolan, G. et al., Haemophilia B: Where are we now and what does the future hold? *Blood reviews*, 2018. 32(1): p. 52-60.
97. Anguela, X.M. et al., Robust ZFN-mediated genome editing in adult hemophilic mice. *Blood*, 2013. 122(19): p. 3283-3287.

Acknowledgments

First of all, I am very grateful to my supervisor Prof. Prof. h.c. mult. Dr. Dr. Bertram Brenig. Thank you for providing me with an impressive and valuable five-year learning experience. You have showed me the important elements in research: solid expertise, rigorous research attitude, patience, optimism and cooperation. At the beginning of my PhD studies, you patiently guided me through all the work that a project needed to go through. As I gained some experience, you gave me enough space and trust to think and explore the various possibilities in scientific research. When I encountered difficulties, I could always get support and encouragement from you. These valuable experiences have allowed me to progress and will have a profound impact on my future scientific work. I would also like to thank you for supporting me to attend conferences and for providing financial support for the final phase of my PhD studies.

I would like to thank my co-supervisor, Prof. Dr. med. Ekkehard Schütz, for providing important suggestions and data support for my research. Many thanks to Prof. Dr. Michael Hölker for his support of this research, and him and Prof. Dr. Jens Tetens for being my graduation defense examiner. Also, I would like to thank all the co-authors in the course of my research, your contributions were critical to this study.

Special thanks go to Dr. rer. nat. Marc Hirschfeld for his advice and experimental guidance throughout the project. You have been like a bridge between me and this lab, which have helped me a lot in the beginning of my PhD studies. And a special thanks to you for proofreading this work. Thanks to Dr. med. vet. Wemheuer for providing me with samples. Also, your stories and care have made my studies here happy and warm. Thanks to Sabrina Pach and Delia Niehaus for their technical supports. Thanks also to Ina Meyer for her kind help in the preparation of the experiments. I would also like to thank Svenja Becker for ELISA and pedigree analysis, Dr. Anne Katrin Hollmann for teaching me Western blot, Dr. Xuying Zhang for helping me get acquainted with the lab, Christin Wacker for sharing her conference experience with me, Lilith Steingraeber and Frederik Krull for their help in daily life. I would like to thank the other members of this lab for supporting me and creating a great working atmosphere for me.

I would like to express my sincere gratitude to the CSC scholarship funded by the China Scholarship Council, which enabled me to study abroad. I would also like to thank Prof. Shuhong Zhao for her consistent support and encouragement. Without your help, none of this would have gone so smoothly.

In addition, I would like to thank my good friends Yao Wang, Mengyu Tu, Zhengcao Li, Xin Wang, Jiliang Li, Cong Tian, Shoukun Zhang, Nana Pang, Yuan Ma, Wanping Chen, Ting Xie etc. for making my life in Göttingen happy and fulfilling. Special thanks to Chunan Guo, who is pursuing her PhD in the Netherlands, for our nearly 20 years of friendship, and for our travels throughout Europe.

My great appreciation goes to my husband, Fangzheng Xu, who is also my research partner. I love the moment when an interesting biological question is posed or an answer slowly surfaces, and I believe you feel the same way. The path of discovery is challenging, but your love and companionship make it easy. I will remember every moment we shared in this beautiful city, whether it was sweet or bitter. You have been the anchor of my PhD journey. Being with you makes me look forward to the future.

Finally, I would like to extend my sincere thanks to my family. Especially my mother, Yuqin Li, and my father, Guanghong Shan, your love is the foundation of everything I have achieved today. Your words have always been with me: "As long as you keep pursuing, one day you will get where you wish to go, even if it may take a long time."

The memory of Göttingen will always remain in my heart, it is so precious and beautiful that it cannot be replaced.

The Chinese acknowledgements are specifically for my family.

为我的家人准备了中文的致谢。

我想向我的家人表示衷心的感谢。你们的养育和教导是我今天所有成就的基础，你们无私又无尽的爱给我成长和前进的勇气。

我有一句座右铭，是母亲送给我的。“只要不停下脚步，总有一天你会到达你想去的地方，尽管这个过程可能很漫长”，这句话包含了很多层面的鼓励，它一直激励着我，脚踏实地，坚持不懈，不必计较一时得失。若渴望发自内心，想做的事情总是可以做到的。如今我已成长为一个可以独立面对这个世界的个体，虽没有栋梁之才，但也小有成就，我想没有辜负你们的期盼。

五年的博士学习充实而美好。我在这里学到了新的知识，感受着人们对待科学问题的不同视角和态度，也结识了很多友善的同事和朋友。我非常幸运的遇到了一位非常好的教授，他在我的博士学习中功不可没。如今我博士毕业了，但这仅仅是科学研究的开始，学习是个终其一生的过程。接下来，我希望能够继续为自然科学的探索 and 进步作出贡献。尽我所能，脚踏实地，追求真理，坚守本心。

哥廷根是一个庄重而美丽的小镇，她大隐隐于市，低调的散发着特有的魅力。生活在这里的人们尊重科学，热爱生活，人们交往饱含热情又相互理解，只有在这里生活过的人才会懂。在哥廷根的日子让我更加热爱生活，也给我更多的动力和热情投入到科研工作中。

最后，再次感谢我的母亲李玉琴，父亲单光洪，姥姥姥爷，爷爷奶奶和老姨。你们都是我尊敬和挚爱的榜样，我爱你们。